GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 49.4422 Seconds
(without alignments)
2917.574 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPNVNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2624.5	92.1	558	16	Q9R396	Q9r396 escherichia
2	2608.5	91.5	558	2	O85506	O85506 escherichia
3	1774.5	62.2	538	2	Q47014	Q47014 escherichia
4	1766.5	62.0	538	2	Q47016	Q47016 escherichia
5	1753.5	61.5	538	2	O85508	O85508 escherichia
6	1570	55.1	551	2	O68258	O68258 escherichia
7	1554.5	54.5	552	2	Q9KWH9	Q9kwh9 escherichia
8	1548.5	54.3	550	2	O52147	O52147 escherichia
9	1522	53.4	547	2	Q9WXK1	Q9wxk1 escherichia
10	1521	53.3	547	2	Q9ETI1	Q9eti1 citrobacter
11	1495	52.4	549	2	O50190	O50190 escherichia
12	190	6.7	1323	5	Q9VI63	Q9vi63 drosophila
13	190	6.7	1376	5	Q9BMZ9	Q9bmz9 drosophila
14	184	6.5	1321	5	Q962D1	Q962d1 drosophila
15	183	6.4	1015	5	Q8SZW9	Q8szw9 drosophila
16	170	6.0	1254	5	Q94185	Q94185 caenorhabdi
17	165.5	5.8	2232	5	Q8IFX6	Q8ifx6 caenorhabdi
18	162.5	5.7	1291	5	O77261	O77261 drosophila
19	162.5	5.7	1300	5	Q9W5E0	Q9w5e0 drosophila
20	161	5.6	1353	5	Q9V8Q2	Q9v8q2 drosophila
21	156	5.5	2921	5	Q9N973	Q9n973 leishmania
22	152.5	5.3	973	16	Q8XDQ4	Q8xdq4 escherichia
23	152.5	5.3	1412	5	Q9VKJ1	Q9vkj1 drosophila
24	149.5	5.2	676	5	Q9VB94	Q9vb94 drosophila
25	149.5	5.2	1241	5	Q8MNI0	Q8mni0 dictyosteli
26	149	5.2	2276	2	Q93TY6	Q93ty6 staphylococ
27	149	5.2	2310	16	Q8CMU7	Q8cmu7 staphylococ
28	148.5	5.2	606	3	Q9P3I9	Q9p3i9 neurospora
29	148	5.2	1192	5	Q9TW45	Q9tw45 caenorhabdi
30	148	5.2	1192	5	Q17346	Q17346 caenorhabdi
31	147.5	5.2	3012	5	O97205	O97205 leishmania
32	145.5	5.1	1550	5	Q9W002	Q9w002 drosophila
33	145	5.1	3848	2	P94772	P94772 erwinia chr
34	145	5.1	4782	11	Q8K1G6	Q8k1g6 mus musculu
35	144.5	5.1	2585	5	Q23587	Q23587 caenorhabdi
36	144	5.1	1323	6	O77509	O77509 bos taurus
37	143.5	5.0	1275	5	O76602	O76602 caenorhabdi
38	143.5	5.0	1805	11	Q63661	Q63661 rattus norv
39	143	5.0	2678	5	Q9NDS4	Q9nds4 dictyosteli
40	143	5.0	3443	11	Q8JZM8	Q8jzm8 mus musculu
41	142.5	5.0	1026	5	Q9V8B9	Q9v8b9 drosophila
42	142.5	5.0	3013	16	Q8ZA73	Q8za73 yersinia pe
43	142	5.0	2283	2	Q8VQ99	Q8vq99 staphylococ
44	141.5	5.0	1027	5	Q9W3U2	Q9w3u2 drosophila
45	141.5	5.0	1096	5	Q17368	Q17368 caenorhabdi

ALIGNMENTS

RESULT 1
Q9R396
ID Q9R396 PRELIMINARY; PRT; 558 AA.
AC Q9R396;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Translocated intimin receptor Tir (Putative translocated intimin
DE receptor protein).
GN TIR OR Z5112 OR ECS4561.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=86/24;
RX MEDLINE=99242825; PubMed=10225900;
RA DeVinney R., Stein M., Reinscheid D., Abe A., Ruschkowski S.,
RA Finlay B.B.;
RT "Enterohemorrhagic Escherichia coli O157:H7 produces Tir, which is
RT translocated to the host cell membrane but is not tyrosine
RT phosphorylated.";
RL Infect. Immun. 67:2389-2398(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=EDL933;
RX MEDLINE=98339885; PubMed=9673266;
RA Perna N.T., Mayhew G.F., Posfai G., Elliott S., Donnenberg M.S.,
RA Kaper J.B., Blattner F.R.;
RT "Molecular evolution of a pathogenicity island from enterohemorrhagic
RT Escherichia coli O157:H7.";
RL Infect. Immun. 66:3810-3817(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

```
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AF125993; AAD29391.1; -.
DR   EMBL; AF071034; AAC31506.1; -.
DR   EMBL; AE005595; AAG58825.1; -.
DR   EMBL; AP002566; BAB37984.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor; Complete proteome.
SQ   SEQUENCE   558 AA;  58022 MW;  99C417222D4B4AA1 CRC64;

  Query Match             92.1%;  Score 2624.5;  DB 16;  Length 558;
  Best Local Similarity   94.0%;  Pred. No. 1e-135;
  Matches  528;  Conservative    3;  Mismatches   24;  Indels    7;  Gaps    3;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60

Qy         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120

Qy        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300

Qy        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360

Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420

Qy        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIG---GPCRIRMLMLKHRCMIRRCRLLILI 477
              ||||||||||||||||||||||||||||||||    |       | | :       :
Db        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIGTVQNPYADVKTSL-HDSQVPTSNSNTSV 479

Qy        478 RLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 537
              :   |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 QNM---GNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 536

Qy        538 GGSNSAVNTSNNPPAPGSHRFV 559
              ||||||||||||||||||||||
Db        537 GGSNSAVNTSNNPPAPGSHRFV 558
```

```
RESULT 2
O85506
ID   O85506      PRELIMINARY;      PRT;   558 AA.
AC   O85506;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Translocated intimin receptor Tir.
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=95SF2;
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
DR   EMBL; AF070067; AAC69314.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   558 AA;  58175 MW;  CA2CDDAC94527C2E CRC64;

  Query Match             91.5%;  Score 2608.5;  DB 2;  Length 558;
  Best Local Similarity   93.4%;  Pred. No. 7.6e-135;
  Matches  525;  Conservative    3;  Mismatches   27;  Indels    7;  Gaps    3;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60

Qy         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120

Qy        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASETETATRDQLTKEAFQNPDNQKVNIDE 300

Qy        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360

Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420

Qy        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIG---GPCRIRMLMLKHRCMIRRCRLLILI 477
              ||||||||||||||||||||||||||||||||    |       | | :       :
Db        421 NVDTPGSEDTMESRRSSMASTLSTFFDTSSIGTVQNPYADVKTSL-HDSQVPTSNSNTSV 479

Qy        478 RLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 537
              :   |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 QNM---GNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALRGGLRHDMGGLT 536

Qy        538 GGSNSAVNTSNNPPAPGSHRFV 559
              ||||||||||||||||||||||
Db        537 GGSNSAVNTSNNPPAPGSHRFV 558
```

```
RESULT 3
Q47014
ID   Q47014      PRELIMINARY;      PRT;   538 AA.
AC   Q47014;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   TRANSLOCATED intimin receptor TIR (EspE protein).
GN   TIR OR ESPE.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REPEC 84/110/1, and E65/56;
RA   Krejany E.O.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=413/89-1;
RX   MEDLINE=98294040; PubMed=9632251;
RA   Deibel C., Kraemer S., Chakraborty T., Ebel F.;
RT   "EspE, a novel secreted protein of attaching and effacing bacteria, is
RT   directly translocated into infected host cells, where it appears as a
RT   tyrosine-phosphorylated 90 kDa protein.";
RL   Mol. Microbiol. 28:463-474(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=95ZG1;
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of Shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10;
RA   Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
RA   Boullier S., De Rycke J., Milon A., Oswald E.;
RT   "Role of Tir and Intimin in the pathogenesis of rabbit
RT   enteropathogenic Escherichia coli.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=413/89-1;
RA   Benkel P., Chakraborty T.;
RT   "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
RT   producing Escherichia coli.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U59502; AAC32028.2; -.
DR   EMBL; AJ223063; CAA11065.1; -.
DR   EMBL; AF070068; AAC69316.1; -.
DR   EMBL; AF132728; AAD27868.1; -.
DR   EMBL; AF113597; AAF03080.1; -.
DR   EMBL; AJ277443; CAC81869.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Hypothetical protein; Receptor.
SQ   SEQUENCE   538 AA;  55420 MW;  31D7A8E227B3D06C CRC64;

  Query Match           62.2%;  Score 1774.5;  DB 2;  Length 538;
  Best Local Similarity  63.2%;  Pred. No. 2.8e-89;
  Matches  360;  Conservative   61;  Mismatches  106;  Indels   43;  Gaps    6;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGG-RGQLINSTGPLGSRALFTPVRNSMADSG 59
              ||||||||||||   |||||||||||||||| | ||||| ||:||| ||||:|||:||:
Db          1 MPIGNLGHNPNVRALIPPAPPLPSQTDGAGGARNQLINSNGPMGSRLLFTPIRNSVADAA 60

Qy         60 DNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHI 119
              |:||||:|||| ||:| ||||::|:   |||| | |||||  ||||:|||||::|| |:
Db         61 DSRASDIPGLPTNPLRFAASEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDGSHV 120

Qy        120 AVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRI 179
              |:||:||:||:||||:||:: |||:|||||:|||||||||||||||||||| ||||||
Db        121 AIGQKNGLETTVVLSEQEPSSLQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEARQRI 180

Qy        180 LELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVA 239
              :: |||| | |:|  |         : |||     |  |||||||||:||||||:|||:|
Db        181 IDKLEPKDTKETKEPG---------DPNSGEGKIIEIHTSTSTSSLRADPKLWLSLGTIA 231

Qy        240 TGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNID 299
              ||||:||||| ||:||||||| | |||||||||:  | | :||||||||||||||||||||
Db        232 AGLIGMAATGIAQAVALTPEPDDPITTDPDAAANTAEAAAKDQLTKEAFQNPDNQKVNID 291

Qy        300 ELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSS 359
              | ||||||| ||||||| | |||||||||:|:|:|||:|:|||:|||||  |||::|: :||
Db        292 ENGNAIPSGELKDDVVAQIAEQAKAAGEQARQEAIESNSQAQQKYDEQHAKREQEMSLSS 351

Qy        360 GAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQ 419
              | |||:|||||||||||  |||||||||| ||| || |         |:|:| ||  ||
Db        352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQTITTRT--------VVDNQPTNNASAQ 403

Qy        420 GNVDTPGSEDTMESRRSSMASTSSTFFDTSS----------IGGPCRIRMLMLKHRCMIR 469
              || || | |::  |||:| || :|   ||||          :| |    :  :    :
Db        404 GNTDTSGPEESPASRRNSNASLASNGSDTSSTGTVENPYADVGMPRNDSLARISEEPIYD 463

Qy        470 RCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGL 529
                          :    || |||   ::   | ||:| |  ||||||| || ||||
Db        464 E-------------VAADPNYSVIQHFSGNSPVTG-RLVGTPGQGIQSTYALLASSGGL 508

Qy        530 RHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |  |||||||  |||:|:|  | ||  |||
Db        509 RLGMGGLTGGGESAVSTANAAPTPGPARFV 538


RESULT 4
Q47016
ID   Q47016      PRELIMINARY;      PRT;   538 AA.
AC   Q47016;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   TRANSLOCATED intimin receptor (Tir).
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REPEC 83/39;
RA   O'Brien R.A.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REPEC 83/39;
RA   Krejany E.O.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RA   Agin T.S., Boedeker E.C.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 172-538 FROM N.A.
RC   STRAIN=REPEC RDEC-1;
RA   Krejany E.O.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-317 FROM N.A.
```

```
RC   STRAIN=RDEC-1;
RX   MEDLINE=97055784; PubMed=8900070;
RA   Agin T.S., Cantey J.R., Boedeker E.C., Wolf M.K.;
RT   "Characterization of the eaeA gene from rabbit enteropathogenic
RT   Escherichia coli strain RDEC-1 and comparison to other eaeA genes from
RT   bacteria that cause attaching-effacing lesions.";
RL   FEMS Microbiol. Lett. 144:249-258(1996).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RX   MEDLINE=98254123; PubMed=9593291;
RA   Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA   Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT   "The complete sequence of the locus of enterocyte effacement (LEE)
RT   from enteropathogenic Escherichia coli E2348/69.";
RL   Mol. Microbiol. 28:1-4(1998).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RX   MEDLINE=21153569; PubMed=11254564;
RA   Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,
RA   Boedeker E.C.;
RT   "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte
RT   Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1.";
RL   Infect. Immun. 69:2107-2115(2001).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RA   Boedeker E.C., Zhu C., Elliott S.J., Tonia T.S., Johnson L.A.,
RA   Thate T.E., Kaper J.B.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=83/39;
RA   Tauschek M., Strugnell R.A., Robins-Browne R.M.;
RT   "Characterization of the LEE pathogenicity islands of rabbit
RT   enteropathogenic Escherichia coli.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U59504; AAD19750.1; -.
DR   EMBL; AF045568; AAC15683.1; -.
DR   EMBL; U59503; AAB02941.1; -.
DR   EMBL; AF200363; AAK26722.1; -.
DR   EMBL; AF453441; AAL57549.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   538 AA;  55410 MW;  3132A969B7B3D06C CRC64;

  Query Match             62.0%;  Score 1766.5;  DB 2;  Length 538;
  Best Local Similarity   63.0%;  Pred. No. 7.8e-89;
  Matches  359;  Conservative   61;  Mismatches  107;  Indels   43;  Gaps    6;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGG-RGQLINSTGPLGSRALFTPVRNSMADSG 59
              |||||||||| ||   |||||||||||||||| | ||||| ||:||| ||||:|||:||:
Db          1 MPIGNLGHNSNVRALIPPAPPLPSQTDGAGGARNQLINSNGPMGSRLLFTPIRNSVADAA 60

Qy         60 DNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHI 119
              |:||||:|||| ||:| ||||::|:   ||||| | |||||  ||||:||||||::|| |:
Db         61 DSRASDIPGLPTNPLRFAASEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDGSHV 120

Qy        120 AVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRI 179
              |:||:||:||:||||:||:: |||:|||||:|||||||||||||||||||| ||||||
Db        121 AIGQKNGLETTVVLSEQEFSSLQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEARQRI 180

Qy        180 LELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVA 239
              :: |||| | |:|  |         : |||      |  |||||||||:||||||:|||:|
Db        181 IDKLEPKDTKETKEPG---------DPNSGEGKIIEIHTSTSTSSLRADPKLWLSLGTIA 231

Qy        240 TGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNID 299
              ||||:||||| ||:|||||| | |||||||:  | | :|||||||||||||||||||||
Db        232 AGLIGMAATGIAQAVALTPEPDDPITTDPDAAANTAEAAAKDQLTKEAFQNPDNQKVNID 291

Qy        300 ELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSS 359
              | ||||||| ||||||| | |||||||||:|:|:|||:|:|||:||||| |||::|: :||
Db        292 ENGNAIPSGELKDDVVAQIAEQAKAAGEQARQEAIESNSQAQQKYDEQHAKREQEMSLSS 351

Qy        360 GAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQ 419
              | |||:|||||||||||  |||||||||| ||| || |          |:|:| ||  ||
Db        352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQTITTRT--------VVDNQPTNNASAQ 403

Qy        420 GNVDTPGSEDTMESRRSSMASTSSTFFDTSS----------IGGPCRIRMLMLKHRCMIR 469
              || || | |::  |||:| || :|   ||||       :| |    :  :   :
Db        404 GNTDTSGPEESPASRRNSNASLASNGSDTSSTGTVENPYADVGMPRNDSLARISEEPIYD 463

Qy        470 RCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGL 529
                          :    || |||   ::   | ||:| |  ||||||| || ||||
Db        464 E-------------VAADPNYSVIQHFSGNSPVTG-RLVGTPGQGIQSTYALLASSGGL 508

Qy        530 RHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |  ||||||  |||:|:|  | ||  |||
Db        509 RLGMGGLTGGGESAVSTANAAPTPGPARFV 538


RESULT 5
O85508
ID   O85508      PRELIMINARY;      PRT;   538 AA.
AC   O85508;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor Tir.
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EPEC87A;
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
DR   EMBL; AF070069; AAC69318.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   538 AA;  55602 MW;  447052A0E3214D6D CRC64;

  Query Match             61.5%;  Score 1753.5;  DB 2;  Length 538;
  Best Local Similarity   62.6%;  Pred. No. 4e-88;
  Matches  357;  Conservative   61;  Mismatches  109;  Indels   43;  Gaps    6;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGG-RGQLINSTGPLGSRALFTPVRNSMADSG 59
              |||||||||||   |||||||||||||||| | ||||| ||:||| ||||:|||:||:
Db          1 MPIGNLGHNPNVRALIPPAPPLPSQTDGAGGARNQLINSNGPMGSRLLFTPIRNSVADAA 60

Qy         60 DNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHI 119
              |:||||:|||| ||:| ||||::|:   ||||| | |||||  ||||:||||||::|| |:
Db         61 DSRASDIPGLPTNPLRFAASEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDGSHV 120

Qy        120 AVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRI 179
              |:||:||:||:|||||||:: :|||:|||||:|||||||||||||||||||| ||||||
Db        121 AIGQKNGLETTVVLSDQEFFSIQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEARQRI 180

Qy        180 LELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVA 239
              :: |||| | |:|  |         : |||      |  ||||| |||:||||||:|||:|
Db        181 IDKLEPKDTKETKEPG---------DPNSGEGKIIEIHTSTSTFSLRADPKLWLSLGTIA 231

Qy        240 TGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNID 299
              ||||:||||| ||:||||||| ||||||| |||  | ||:|:|||:||||:|| ||||||
```

```
Db        232 AGLIGMAATGIAQAVALTPEPDDPTTTDPDTAASTAEAATKDRLTQEAFQDPDKQKVNID 291

Qy        300 ELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSS 359
              | ||||||| | ||||| | |||||||||:|:|:|||:|:||||||||| |||::|: :||
Db        292 ENGNAIPSGELIDDVVAQIAEQAKAAGEQARQEAIESNSQAQKKYDEQHAKREQEMSLSS 351

Qy        360 GAGYGLSGALILGGGIGAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQ 419
              | |||:|||||||||||  |||||||||| ||| || |          |:|:| ||  ||
Db        352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQTITTRT--------VVDNQPTNNASAQ 403

Qy        420 GNVDTPGSEDTMESRRSSMASTSSTFFDTSS----------IGGPCRIRMLMLKHRCMIR 469
              || || | |::  |||:| || :|   ||||          :| |    :  :    :
Db        404 GNTDTSGPEESPASRRNSNASLASNGSDTSSTGTVENPYADVGMPRNDSLARIPEEPIYD 463

Qy        470 RCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGL 529
                          :     || |||   ::   | ||:| |  ||||||| || ||||
Db        464 E-------------VAADPNYSVIQHFSGNSPVTG-RLVGTPGQGIQSTYALLASSGGL 508

Qy        530 RHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |  ||||||  |||:|:|  | ||  |||
Db        509 RLGMGGLTGGGESAVSTANASPTPGPARFV 538


RESULT 6
O68258
ID   O68258      PRELIMINARY;      PRT;   551 AA.
AC   O68258;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98187918; PubMed=9529069;
RA   Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT   "Molecular analysis of Shiga toxigenic Escherichia coli O111:H-
RT   proteins which react with sera from patients with hemolytic-uremic
RT   syndrome.";
RL   Infect. Immun. 66:1467-1472(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
DR   EMBL; AF025311; AAC69249.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   551 AA;  56975 MW;  6EC95F76EF0F44CC CRC64;

  Query Match             55.1%;  Score 1570;  DB 2;  Length 551;
  Best Local Similarity   59.5%;  Pred. No. 4.4e-78;
  Matches  339;  Conservative   67;  Mismatches  134;  Indels   30;  Gaps   11;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
              |||||||:| | || ||||||||||||||   || |||||||| |||| ||:|:|||:|||
Db          1 MPIGNLGNNVNSNNLIPPAPPLPSQTDGASRGGAGQLINSTGALGSRLLFSPLRNSIADS 60

Qy         59 GDNRASDVPGLPVNPMRL--AASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116
                |:|  |:|||||:| ||  | ||| |  ||||||| |||||||:|||:| ||:| | ||
Db         61 VDSR--DIPGLPVHPSRLATATSEICLLGGFEVLHDKGPLDTLNKQIGASAFRIEQQSDG 118

Qy        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176
                : |:|::|||| ||:|: ||   ||:|| | | :||||||||| ||:||| |||| |||

Db        119 SYAAIGEKNGVEVSVILNSQELQSLQAIDIEDKGRFVFTGGRGGGGHSMVTPASDIAEAR 178

Qy        177 QRILELLEPKGTGESKGAG---ESKGVGELRESNSGAENT--TETQTSTSTSSLRSDPKL 231
                :||  |:|   | |:        | |||     | || :::  :||:||:: ||:|||||
Db        179 AKILAKLDPNNHGGSQARNVDTRSVGVG----SASGMDDSVVSETRTSSTASSVRSDPKF 234

Qy        232 WLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNP 291
              |:::| :| || ||||||| |||||||||| ||||||: |||| |:|||||||:|||:||
Db        235 WVSIGAIAAGLAGLAATGITQALALTPEPDDPTTTDPEQAASAAESATRDQLTQEAFKNP 294

Qy        292 DNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              :||||:|||:||:|||| ||||||| |||||| ||| |:|||:|:||||::|| | |:|
Db        295 ENQKVSIDEIGNSIPSGELKDDVVAKIEEQAKEAGEAARQQAVESNAQAQQRYDTQYARR 354

Qy        352 QEELKVSSGAGYGLSGALILGGGIGAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENK 411
              ||||::||| || || |||:|||||  || ||||:||| |||||||| |        |: :
Db        355 QEELELSSGIGYSLSSALIVGGGIGAGVTTALHRRNQPAEQTTTTTTHT------VVQQQ 408

Qy        412 PANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFF-DTSSIGGPCRIRMLMLKHRCMIRR 470
              |||||| |   :|||  :|| |  ||:|| : ||||        :   ::    |:
Db        409 TGGNTPAQGGTDAIRAEDTSLNRRDSQRSTASTHWSDTSSAVVNPYAEVGEARNSSPARQ 468

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA-LSGGL 529
                  |          :     || ||:   :    | ||:| |  ||||||| |    | ||
Db        469 AEEHIYDE------VAADPNYSVIQNFSGNNQVTG-RLMGTPGQGIQSTYAILTNNSAGL 521

Qy        530 RHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |  ||||||   |||||:|  | ||  |||
Db        522 RLGMGGLTGSGGSAVNTANAAPTPGPGRFV 551


RESULT 7
Q9KWH9
ID   Q9KWH9      PRELIMINARY;      PRT;   552 AA.
AC   Q9KWH9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Tir.
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK01;
RA   Abe A., Nagano H.;
RT   "Analyses of type III secreted proteins and Tir in enteropathogenic
RT   Escherichia coli O157:H45.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB036053; BAA96815.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
SQ   SEQUENCE   552 AA;  57005 MW;  A8D79EE22EE50A4B CRC64;

  Query Match             54.5%;  Score 1554.5;  DB 2;  Length 552;
  Best Local Similarity   57.5%;  Pred. No. 3.1e-77;
  Matches  336;  Conservative   63;  Mismatches  128;  Indels   57;  Gaps   12;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
              |||||||:| | |: |||||||||||||  || | ||:||| |||||::|:|:|||:|||
Db          1 MPIGNLGNNVNSNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSVADS 60

Qy         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116
                |:|  |:|||| || ||||  ||  |  ||||||| |||||||:||| | |||| | ||
Db         61 VDSR--DIPGLPANPSRLAAATSETCLLGGFEVLHDKGPLDTLNQQIGPSAFRVEAQADG 118

Qy        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176
                | |:|::||:| || || ||   ||||| |||::|||||||||:|| |||||||| |||
Db        119 THAAIGEKNGLEVSVALSPQELQSLQSIDIEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
```

```
Qy        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227
Db        179 MKILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230

Qy        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287
Db        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290

Qy        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347
Db        291 FKNPENQKVNIDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350

Qy        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSART 407
Db        351 YARRQEELQLSSGIGYGLSSALIGAGGIGAGVTTALHRRNQPAEQTTTTTTHT------V 404

Qy        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS---------IGGPCRIR 458
Db        405 VQQQTGGNTPAQGGTDATRAEDASLNRRDSQGSVASTHWSDSSSEVVNPYAEVGEP---- 460

Qy        459 MLMLKHRCMIRRCRLLILIRLFRIWG-IQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQ 517
Db        461 -----------RNSLSTRQQEEHIYDEVAADPVYSVIQNFSRNAPVTG-RLMGSPGQGIQ 508

Qy        518 STYARLALSGGLRHDMGGLTGGSNSAVNT--SNNPPAPGSHRFV 559
Db        509 STYALLANSAGLRLGMGGLTGSGESAVNTANANAAPTPGPVRFV 552
```

RESULT 8
O52147
ID O52147 PRELIMINARY; PRT; 550 AA.
AC O52147;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tir.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98254123; PubMed=9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
RT from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
DR EMBL; AF022236; AAC38390.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
SQ SEQUENCE 550 AA; 56509 MW; 19DD08A9BE9251CB CRC64;

Query Match 54.3%; Score 1548.5; DB 2; Length 550;
Best Local Similarity 56.0%; Pred. No. 6.6e-77;
Matches 330; Conservative 67; Mismatches 123; Indels 69; Gaps 12;

```
Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
Db          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

Qy         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116
Db         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118

Qy        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176
Db        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178

Qy        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227
Db        179 TKILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230

Qy        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287
Db        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290

Qy        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347
Db        291 FKNPENQKVNIDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350

Qy        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSART 407
Db        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHT------V 404

Qy        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS---------IGGPCRIR 458
Db        405 VQQQTGGNTPAQGGTDATRAEDASLNRRDSQGSVASTHWSDSSSEVVNPYAEVGG---AR 461

Qy        459 MLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPP-----RDTTDNG---ARLLG 510
Db        462 NSLSAHQ--------------------PEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIG 501

Qy        511 NPSAGIQSTYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
Db        502 TPGQGIQSTYALLANSGGLRLGMGGLTSGGESAVSSVNAAPTPGPVRFV 550
```

RESULT 9
Q9WXK1
ID Q9WXK1 PRELIMINARY; PRT; 547 AA.
AC Q9WXK1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor (Translocated intimin receptor
DE Tir).
GN TIR.
OS Escherichia coli, and
OS Citrobacter rodentium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 67825;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=MPEC;
RA Okutani A., Itoh K., Sasakawa C.;
RT "Translocated intimin receptor(Tir) of murine pathogenic Escherichia
RT coli O115a,c:K(B).";
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.rodentium; STRAIN=DBS100;
RX MEDLINE=21437640; PubMed=11553577;
RA Deng W., Li Y., Vallance B.A., Finlay B.B.;
RT "Locus of Enterocyte Effacement from Citrobacter rodentium: Sequence
RT Analysis and Evidence for Horizontal Transfer among Attaching and
RT Effacing Pathogens.";
RL Infect. Immun. 69:6323-6335(2001).
DR EMBL; AB026719; BAA77400.1; -.
DR EMBL; AF311901; AAL06376.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
KW Receptor.
SQ SEQUENCE 547 AA; 56283 MW; 02CAC6D625FA6EE1 CRC64;

Query Match 53.4%; Score 1522; DB 2; Length 547;
Best Local Similarity 57.2%; Pred. No. 1.8e-75;
Matches 327; Conservative 68; Mismatches 139; Indels 38; Gaps 10;

```
Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
```

```
Db          1 MPIGNLGNNNISNNLIPPAPPLPSQTDGATRGNGSSLISSTGSLGSRLLFSPLRSSIVDT 60

Qy         59 GDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKH 118
              |:|  ||||||  :|:| | ||  |: |||||||  |||||||::|||||||| | || |
Db         61 VDSR--DVPGLPEHPLRFATSETCLHGGFEVLHDKGPLDTLNQKIGSSVFRVEQQPDGTH 118

Qy        119 IAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQR 178
              |:| ::||| || |:  |   |||:| ||  :|||||||||:||||||||||::||::
Db        119 AAIGVKDGVEVSVTLNSSELQSLQSLDTEGNGRFVFTGGRGGSGHAMVTVASDISQAREK 178

Qy        179 ILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDP 229
              |:  |:|   |           | | | : |:|:       |    :|| |||:|||:||||
Db        179 IIAKLDPDNHGGRQPKDIDTRSVGVGSASGMGD------GV--VSETHTSTTTSSVRSDP 230

Qy        230 KLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQ 289
              | |:::| :| || ||||||||||:|||| || ||||||| ||:| | ||:|||||||||
Db        231 KFWVSVGAIAAGLAGLAATGIVQAVALTPAPDDPTTTDPDEAANAAEAATKDQLTKEAFQ 290

Qy        290 NPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQA 349
              |||||||||||||||||| ||||||| | :||| |||:|:|||:|:|||||:::|:|||
Db        291 NPDNQKVNIDELGNAIPSGELKDDVVAQIADQAKVAGEQARQQAVESNAQAQQRHDDQQA 350

Qy        350 KRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVE 409
              |||:|| :||| ||||| |||:|||||  ||| |||:| | |||  || :        ::
Db        351 KRQQELDLSSGIGYGLSSALIVGGGIGAGVTAMLHRRNPPTEQTIATTHS-------VIQ 403

Qy        410 NKPANNTPAQGNVDTPGSEDTMESRRSSMASTSST-FFDTS-SIGGPCRIRMLMLKHRCM 467
              :   || |||  || | |:   :|| | || :|| : |||  :  |                |
Db        404 QQTGGNTRAQGGADTTGVENASLTRRDSQASVASTQWSDTSGDVVNP-------YAEGWM 456

Qy        468 IRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSG 527
              |    |:          :     || |||   :    | ||:|:|  |||||| || ||
Db        457 SRNNPSLLAPEEPIYDEVAPDPNYSVIQHFSGNNPVTG-RLVGSPGQGIQSTYALLASSG 515

Qy        528 GLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |||  |||||||  || : :|    ||  |||
Db        516 GLRLGMGGLTGGGESAGSAANAATTPGVERFV 547


RESULT 10
Q9ETI1
ID   Q9ETI1      PRELIMINARY;      PRT;   547 AA.
AC   Q9ETI1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor Tir.
OS   Citrobacter rodentium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Citrobacter.
OX   NCBI_TaxID=67825;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1843-73T, and DBS100;
RX   MEDLINE=20553330; PubMed=11101562;
RA   Luperchio S.A., Newman J.V., Dangler C.A., Schrenzel M.D.,
RA   Brenner D.J., Steigerwalt A.G., Schauer D.B.;
RT   "Citrobacter rodentium, the Causative Agent of Transmissible Murine
RT   Colonic Hyperplasia, Exhibits Clonality: Synonymy of C. rodentium and
RT   Mouse-Pathogenic Escherichia coli.";
RL   J. Clin. Microbiol. 38:4343-4350(2000).
DR   EMBL; AF301618; AAG40758.1; -.
DR   EMBL; AF301617; AAG25642.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   547 AA;  56270 MW;  CB8318B301049C37 CRC64;

  Query Match           53.3%;  Score 1521;  DB 2;  Length 547;
  Best Local Similarity   57.2%;  Pred. No. 2.1e-75;
  Matches  327;  Conservative   68;  Mismatches  139;  Indels   38;  Gaps   10;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
              |||||||:|   || |||||||||||||  |    ||:||| |||| ||:|:|:|: |:
Db          1 MPIGNLGNNNISNNLIPPAPPLPSQTDGATRGNGSSLISSTGSLGSRLLFSPLRSSIVDT 60

Qy         59 GDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKH 118
              |:|  |||||| :|:| | ||  |: ||||||| |||||||::|||||||| | || |
Db         61 VDSR--DVPGLPEHPLRFATSETCLHGGFEVLHDKGPLDTLNQKIGSSVFRVEQQPDGTH 118

Qy        119 IAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQR 178
              |:| ::||| || |:  |   |||:| ||  :|||||||||:||||||||||::||::
Db        119 AAIGVKDGVEVSVTLNSSELQSLQSLDTEGTGRFVFTGGRGGSGHAMVTVASDISQAREK 178

Qy        179 ILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDP 229
              |:  |:|   |           | | | : |:|:       |    :|| |||:|||:||||
Db        179 IIAKLDPDNHGGRQPKDIDTRSVGVGSASGMGD------GV--VSETHTSTTTSSVRSDP 230

Qy        230 KLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQ 289
              | |:::| :| || ||||||||||:|||| || ||||||| ||:| | ||:|||||||||
Db        231 KFWVSVGAIAAGLAGLAATGIVQAVALTPAPDDPTTTDPDEAANAAEAATKDQLTKEAFQ 290

Qy        290 NPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQA 349
              |||||||||||||||||| ||||||| | :||| |||:|:|||:|:|||||:::|:|||
Db        291 NPDNQKVNIDELGNAIPSGELKDDVVAQIADQAKVAGEQARQQAVESNAQAQQRHDDQQA 350

Qy        350 KRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVE 409
              |||:|| :||| ||||| |||:|||||  ||| |||:| | |||  || :        ::
Db        351 KRQQELDLSSGIGYGLSSALIVGGGIGAGVTAMLHRRNPPTEQTIATTHS-------VIQ 403

Qy        410 NKPANNTPAQGNVDTPGSEDTMESRRSSMASTSST-FFDTS-SIGGPCRIRMLMLKHRCM 467
              :   || |||  || | |:   :|| | || :|| : |||  :  |                |
Db        404 QQTGGNTRAQGGADTTGVENASLTRRDSQASVASTQWSDTSGDVVNP-------YAEGWM 456

Qy        468 IRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSG 527
              |    |:          :     || |||   :    | ||:|:|  |||||| || ||
Db        457 SRNNPSLLAPEEPIYDEVAPDPNYSVIQHFSGNNPVTG-RLVGSPGQGIQSTYALLASSG 515

Qy        528 GLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |||  |||||||  || : :|    ||  |||
Db        516 GLRLGMGGLTGGGESAGSAANAATTPGVERFV 547


RESULT 11
O50190
ID   O50190      PRELIMINARY;      PRT;   549 AA.
AC   O50190;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor.
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=E2348/69;
RX   MEDLINE=98050926; PubMed=9390560;
RA   Kenny B., DeVinney R., Stein M., Reinscheid D.J., Frey E.A.,
RA   Finlay B.B.;
RT   "Enteropathogenic E. coli (EPEC) transfers its receptor for intimate
RT   adherence into mammalian cells.";
RL   Cell 91:511-520(1997).
DR   EMBL; AF013122; AAB88410.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   549 AA;  56843 MW;  40C8B8B234409A08 CRC64;

  Query Match           52.4%;  Score 1495;  DB 2;  Length 549;
  Best Local Similarity   54.7%;  Pred. No. 5.5e-74;
  Matches  322;  Conservative   65;  Mismatches  132;  Indels   70;  Gaps   12;
```

```
Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
              ||||||||:| | |: ||||||||||||||  || | ||:||| ||||:||:|:|||||||
Db          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

Qy         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116
                |:|  |:|||| || ||||  ||  |  ||||||| |||| || ||| | |||| | ||
Db         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118

Qy        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176
              | |:|::||:| || || ||:: ||||| |||::|||||||||:|| |||||||| |||
Db        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178

Qy        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227
              :||  |:|   |          | | | : |:      : |    :|| |||: ||:||
Db        179 TKILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230

Qy        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287
              ||| |:::| :| || ||||||| |||||||||| ||||||| ||:| |:||:||||:||
Db        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290

Qy        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347
              |:||:|||||||  ||||||| ||||:|  | :||| ||| |:|||:|:|||||::|::|
Db        291 FKNPENQKVNIDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350

Qy        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSART 407
              |:|||||::||| ||||| |||: |||| || ||||:||| ||||||||          |
Db        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403

Qy        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS---------IGGPCRIR 458
              |  :    |       |         || |  | :|| :  ||         :||    |
Db        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGG---AR 460

Qy        459 MLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPP-----RDTTDNG---ARLLG 510
              :  |:                          :|  :   |     :: : :|    ||:|
Db        461 NSLSAHQ-------------------PEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIG 500

Qy        511 NPSAGIQSTYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |  |||||||| || |||||  ||||| |  :||:: |  | ||  |||
Db        501 TPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549


RESULT 12
Q9VI63
ID   Q9VI63      PRELIMINARY;      PRT;  1323 AA.
AC   Q9VI63; Q9NHX4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG1070 protein (Alhambra).
GN   ALHAMBRA OR CG1070.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Perrin L., Dura J.M.;
RT   "Alhambra, a Drosophila homolog of mammalian AF-10 and AF-17.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003672; AAF54062.2; -.
DR   EMBL; AF217960; AAF72595.1; -.
DR   FlyBase; FBgn0037471; Alhambra.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   PRINTS; PR00308; ANTIFREEZEI.
```

```
DR   SMART; SM00249; PHD; 2.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   1323 AA;  132884 MW;  D53C0C8AF392F9A6 CRC64;

  Query Match             6.7%;  Score 190;  DB 5;  Length 1323;
  Best Local Similarity   20.2%;  Pred. No. 0.038;
  Matches  131;  Conservative   90;  Mismatches  273;  Indels  154;  Gaps   23;

Qy          3 IGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNR 62
              |||: :  ::||::|      | |  ||          | ||  : :        | ::
Db        322 IGNISN--SLNNNLPGG---SSSTSSAG--------NVPGGSGGIISASSGGATQSTSSQ 368

Qy         63 ASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVG 122
              :|  ||   :    |:|  : :: ::  |      :|::   |        :||| :
Db        369 SSTAPGTTKSS---ASSSSSSSNSYKEKHS----KSLSKSTSSK------DKDGKDSSTN 415

Qy        123 QRNGVET----SVVLSDQEYARLQSIDPEGKDK-------------------------F 152
                |   |    |   |:    | :|     |:|
Db        416 SANNNFTNSSASSTSSNSSSTREKSSSKLSKNKDSNQVPSATSSLSTTSSINTQPSSSTS 475

Qy        153 VFTGGRGGAG-HAMVTVASDITEARQRILE------LLEPKGTGESKGAGESKGVGELRE 205
              | | || | |   : || :  |     |        |   |||     ||: : |  | ||
Db        476 TATAGSGGTGTHVSSSAASGLNSAPSTTNEHSNHAHTLSTNGTGAGSAAGKLQSVSNLSN 535

Qy        206 SNSG-----------AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAA--TGIVQ 252
              |:||           : :|     |     |:  |: :     |: |: : |  |  || |
Db        536 SSSGFGSDLRSVSTSSSSTVNDSTGGFGSNSNSERENLSGAGSSASNMPGTIAPGTGGVS 595

Qy        253 ALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKD 312
              : | |    :   :    | | | |:|    :  :  |  ::|   |
Db        596 SSAATNLSTNKGGSSSSTANSLTSTST----SSGSSSNSSSKKRKAD------------ 638

Qy        313 DVVANIEEQAKAAGE-EAKQQAIENNAQAQKKYD-----------------EQQAKRQE 353
                  ||:          |:|:|     :||                  |: :|||
Db        639 --------SAKSTSSISTSGSALEDNNSLISRYDIKDVQVALTPLTDFEKEIEKSSKRQR 690

Qy        354 -ELKV---SSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVE 409
               ||      : |   ::  |       :||||:    : |    |||  :: : :| :
Db        691 TELSPPTHQTSATAEVNAPLASSTSASIAVTASATAASAPPATGTTTLGSSISGNAGSTS 750

Qy        410 NKPANNTPAQGNVDT-------PGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLML 462
              :  ::   | |   :       | :| :  |  :|  | | :  :|||       |: :
Db        751 SGNSSGVSATGGAQSAVGSGGYPKTESSKSSGTASAGSGSGSSSNTSSTKHGSNIKDISS 810

Qy        463 KHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHP----PRDTTDNG----ARLLGNPSA 514
                          |:      :||  ||   |   | :| :     :    : |:
Db        811 SSNQQASTASSSSAPSLY------VSVPLSTANVPGINLPTSSTSSSTTSESHSASSRSS 864

Qy        515 GIQSTYARLALSGGLRHDMGGLTG---------GSNSAVNTSNNPPAP 553
              | || : :   :  :  |||  |          ||:| :   :   :|
Db        865 GAQSQHQQQLSNALVGPPMGGSAGAFHGGTTSAGSSSVIQQQSGKSSP 912


RESULT 13
Q9BMZ9
ID   Q9BMZ9      PRELIMINARY;      PRT;  1376 AA.
AC   Q9BMZ9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   AF10.
GN   ALHAMBRA OR CG1070.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Linder B., Jackle H.;
RT   "The Drosophila homolog of AF10 is expressed in a striped pattern
RT   throughout blastoderm.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AF281145; AAK06385.1; -.
DR   FlyBase; FBgn0037471; Alhambra.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 2.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   1376 AA;  138906 MW;  A40DF3CF5121B17C CRC64;

  Query Match             6.7%;  Score 190;  DB 5;  Length 1376;
  Best Local Similarity   20.2%;  Pred. No. 0.04;
  Matches  131;  Conservative   90;  Mismatches  273;  Indels  154;  Gaps   23;

Qy          3 IGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNR 62
              |||: :  ::||::|      | |  ||          | ||  : :        | ::
Db        375 IGNISN--SLNNNLPGG---SSSTSSAG--------NVPGGSGGIISASSGGATQSTSSQ 421

Qy         63 ASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVG 122
              :|  ||   :    |:|  : :: ::  |      :|::   |        :||| :
Db        422 SSTAPGTTKSS---ASSSSSSSNSYKEKHS----KSLSKSTSSK------DKDGKDSSTN 468

Qy        123 QRNGVET----SVVLSDQEYARLQSIDPEGKDK-------------------------F 152
                |   |    |   |:    | :|     |:|
Db        469 SANNNFTNSSASSTSSNSSSTREKSSSKLSKNKDSNQVPSATSSLSTTSSINTQPSSSTS 528

Qy        153 VFTGGRGGAG-HAMVTVASDITEARQRILE------LLEPKGTGESKGAGESKGVGELRE 205
              | | || | |   : || :  |     |        |   |||     ||: : |  | ||
Db        529 TATAGSGGTGTHVSSSAASGLNSAPSTTNEHSNHAHTLSTNGTGAGSAAGKLQSVSNLSN 588

Qy        206 SNSG-----------AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAA--TGIVQ 252
              |:||           : :|     |     |:  |: :     |: |: : |  |  || |
Db        589 SSSGFGSDLRSVSTSSSSTVNDSTGGFGSNSNSERENLSGAGSSASNMPGTIAPGTGGVS 648

Qy        253 ALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKD 312
              : | |    :   :    | | | |:|    :  :  |  ::|   |
Db        649 SSAATNLSTNKGGSSSSTANSLTSTST----SSGSSSNSSSKKRKAD------------ 691

Qy        313 DVVANIEEQAKAAGE-EAKQQAIENNAQAQKKYD-----------------EQQAKRQE 353
                  ||:          |:|:|     :||                  |: :|||
Db        692 --------SAKSTSSISTSGSALEDNNSLISRYDIKDVQVALTPLTDFEKEIEKSSKRQR 743

Qy        354 -ELKV---SSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVE 409
               ||      : |   ::  |       :||||:    : |    |||  :: : :| :
Db        744 TELSPPTHQTSATAEVNAPLASSTSASIAVTASATAASAPPATGTTTLGSSISGNAGSTS 803

Qy        410 NKPANNTPAQGNVDT-------PGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLML 462
              :  ::   | |   :       | :| :  |  :|  | | :  :|||       |: :
Db        804 SGNSSGVSATGGAQSAVGSGGYPKTESSKSSGTASAGSGSGSSSNTSSTKHGSNIKDISS 863

Qy        463 KHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHP----PRDTTDNG----ARLLGNPSA 514
                          |:      :||  ||   |   | :| :     :    : |:
Db        864 SSNQQASTASSSSAPSLY------VSVPLSTANVPGINLPTSSTSSSTTSESHSASSRSS 917

Qy        515 GIQSTYARLALSGGLRHDMGGLTG---------GSNSAVNTSNNPPAP 553
              | || : :   :  :  |||  |          ||:| :   :   :|
Db        918 GAQSQHQQQLSNALVGPPMGGSAGAFHGGTTSAGSSSVIQQQSGKSSP 965


RESULT 14
Q962D1
ID   Q962D1      PRELIMINARY;      PRT;  1321 AA.
AC   Q962D1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zinc finger/leucine zipper protein DALF isoform C3.
GN   ALHAMBRA OR CG1070.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=21095180; PubMed=11165485;
RA   Bahri S.M., Chia W., Yang X.;
RT   "The Drosophila homolog of human AF10/AF17 leukemia fusion genes
RT   (Dalf) encodes a zinc finger/leucine zipper nuclear protein required
RT   in the nervous system for maintaining EVE expression and normal
RT   growth.";
RL   Mech. Dev. 100:291-301(2001).
DR   EMBL; AY036105; AAK63171.1; -.
DR   FlyBase; FBgn0037471; Alhambra.
DR   InterPro; IPR006895; zf-Sec23_Sec24.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF04810; zf-Sec23_Sec24; 1.
DR   SMART; SM00249; PHD; 2.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   1321 AA;  132749 MW;  3E7ECC724CB19AF7 CRC64;

  Query Match             6.5%;  Score 184;  DB 5;  Length 1321;
  Best Local Similarity   20.0%;  Pred. No. 0.081;
  Matches  130;  Conservative   92;  Mismatches  269;  Indels  160;  Gaps   24;

Qy          3 IGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNR 62
              |||: :  ::||::|      | |  ||          | ||  : :        |  ::
Db        321 IGNISN--SLNNNLPGG---SSSTSSAG--------NVPGGSGGIISASSGGATQSTSSQ 367

Qy         63 ASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVG 122
              :|  ||   :    |:|  : :: ::  |      :|::   |        :|||  :
Db        368 SSTAPGTTKSS---ASSSSSSSNSYKEKHS----KSLSKSTSSK------DKDGKDSSTN 414

Qy        123 QRNGVETSVVLSD---------QEYARLQSIDPEGKDK---------------------- 151
                 |   |:   |          ::|:   |   : ||
Db        415 SANNNFTNSSASSTSSNSSTTREKYSSKLS---KNKDSNQVPSATSSLSTTSSINTQPSS 471

Qy        152 --FVFTGGRGGAG-HAMVTVASDITEARQRILE------LLEPKGTGESKGAGESKGVGE 202
                 | || | |   : || :  |       |       |   |||    ||: : |
Db        472 STSTATAGSGGTGTHVSSSAASGLNSAPSTTNEHSNHAHTLSTNGTGAGSAAGKLQSVSN 531

Qy        203 LRESNSG-----------AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAA--TG 249
              |  |:||           : :|     |     |:  |: :      |: |: : |  |  ||
Db        532 LSNSSSGFGSDLRSVSTSSSSTVNDSTGGFGSNSNSERENLSGAGSSASNMPGTIAPGTG 591

Qy        250 IVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGV 309
               | : | |    :   :    | | | |:|    :  :  |  ::|   |
Db        592 GVSSSAATNLSTNKGGSSSSTANSLTSTST----SSGSSSNSSSKKRKAD---------- 637

Qy        310 LKDDVVANIEEQAKAAGE-EAKQQAIENNAQAQKKYD-----------------EQQAK 350
                          ||:          |:|:|     :||               |: :|
Db        638 ----------SAKSTSSISTSGSALEDNNSLISRYDIKDVQVALTPLTDFEKEIEKSSK 686

Qy        351 RQE-ELKV---SSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSAR 406
              ||  ||       : |   ::  |       :||||:    : |    |||  :: : :|
Db        687 RQRTELSPPTHQTSATAEVNAPLASSTSASIAVTASATAASAPPATGTTTLGSSISGNAG 746

Qy        407 TVENKPANNTPAQGNVDT-------PGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRM 459
              :  :  ::   | |   :       | :| :  |  :|  | | :  :|||       |:
Db        747 STSSGNSSGVSATGGAQSAVGSGGYPKTESSKSSGTASAGSGSGSSSNTSSTKHGSNIKD 806

Qy        460 LMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHP----PRDTTDNG----ARLLGN 511
              :                  |:      :||  ||    |    | :| :     :    :
Db        807 ISSSSNQQASTASSSSAPSLY------VSVPLSTANVPGINLPTSSTSSSTTSESHSASS 860

Qy        512 PSAGIQSTYARLALSGGLRHDMGGLTG---------GSNSAVNTSNNPPAP 553
               |:| || : :   :  :   |||  |          ||:| :   :   :|
Db        861 RSSGAQSQHQQQLSNALVGPPMGGSAGAFHGGTTSAGSSSVIQQQSGKSSP 911


RESULT 15
Q8SZW9
ID   Q8SZW9      PRELIMINARY;      PRT;  1015 AA.
AC   Q8SZW9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   LP05501p.
GN   ALHAMBRA OR CG1070.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY069737; AAL39882.1; -.
DR   FlyBase; FBgn0037471; Alhambra.
SQ   SEQUENCE   1015 AA;  100464 MW;  CB24616CF13D4C53 CRC64;

  Query Match             6.4%;  Score 183;  DB 5;  Length 1015;
  Best Local Similarity   20.1%;  Pred. No. 0.065;
  Matches  130;  Conservative   89;  Mismatches  275;  Indels  154;  Gaps   23;

Qy          3 IGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNR 62
              |||: :  ::||::|      | |  ||          | ||  : :        |  ::
Db         14 IGNISN--SLNNNLPGG---SSSTSSAG--------NVPGGSGGIISASSGGATQSTSSQ 60

Qy         63 ASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVG 122
              :|  ||   :    |:|  : :: ::  |      :|::   |        :|||  :
Db         61 SSTAPGTTKSS---ASSSSSSSNSYKEKHS----KSLSKSTSSK------DKDGKDSSTN 107

Qy        123 QRNGVET----SVVLSDQEYARLQSIDPEGKDK-------------------------F 152
                 |    |    |   |:    | :|      :|
Db        108 SANNNFTNSSASSTSSNSSSTREKSSSKLSNNKDSNQVPSATSSLSTTSSINTQPSSSTS 167

Qy        153 VFTGGRGGAG-HAMVTVASDITEARQRILE------LLEPKGTGESKGAGESKGVGELRE 205
                | | || | |   : || :  |       |       |   |||    ||: : |  |
Db        168 TATAGSGGTGTHVSSSAASGLNSAPSTTNEHSNHAHTLSTNGTGAGSAAGKLQSVSNLSN 227

Qy        206 SNSG-----------AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAA--TGIVQ 252
              |:||           : :|     |     |:  |: :      |: |: : |  |  || |
Db        228 SSSGFGSDLRSVSTSSSSTVNDSTGGFGSNSNSERENLSGAGSSASNMPGTIAPGTGGVS 287

Qy        253 ALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKD 312
              : | |    :   :    | | | | |    :  :  |  ::|   |
Db        288 SSAATNLSTNKGGSSSSTANSLTSTIT----SSGSSSNSSSKKRKAD------------- 330

Qy        313 DVVANIEEQAKAAGE-EAKQQAIENNAQAQKKYD-----------------EQQAKRQE 353
                          ||:          |:|:|     :||               |: :|||
Db        331 --------SAKSTSSISTSGSALEDNNSLISRYDIKDVQVALTPLTDFEKEIEKSSKRQR 382

Qy        354 -ELKV---SSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVE 409
               ||       : |   ::  |       :||||:    : |    |||  :: : :| :
Db        383 TELSPPTHQTSATAEVNAPLASSTSASIAVTASATAASAPPATGTTTLGSSISGNAGSTS 442

Qy        410 NKPANNTPAQGNVDT-------PGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLML 462
              :  ::   | |   :       | :| :  |  :|  | | :  :|||       |: :
Db        443 SGNSSGVSATGGAQSAVGSGGYPKTESSKSSGTASAGSGSGSSSNTSSTKHGSNIKDISS 502

Qy        463 KHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHP----PRDTTDNG----ARLLGNPSA 514
                           |:      :||  ||    |    | :| :     :    : |:
Db        503 SSNQQASTASSSSAPSLY------VSVPLSTANVPGINLPTSSTSSSTTSESHSASSRSS 556

Qy        515 GIQSTYARLALSGGLRHDMGGLTG---------GSNSAVNTSNNPPAP 553
               | || : :   :  :   |||  |          ||:| :   :   :|
Db        557 GAQSQHQQQLSNALVGPPMGGSAGAFHGGTTSAGSSSVIQQQSGKSSP 604


Search completed: December 16, 2003, 08:55:45
Job time : 52.4422 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 21.694 Seconds
(without alignments)
1211.758 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPNVNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	146.5	5.1	1120	1	STFR_ECOLI	P76072 escherichia
2	142	5.0	563	1	MUCS_BOVIN	P98091 bos taurus
3	140.5	4.9	1026	1	STAU_DROME	P25159 drosophila
4	136.5	4.8	461	1	US45_LACLC	P22865 lactococcus
5	133.5	4.7	3178	1	YS89_CAEEL	Q09624 caenorhabdi
6	133	4.7	439	1	Y579_CHLTR	O84583 chlamydia t
7	132.5	4.6	755	1	Y572_CHLPN	Q9z7y1 chlamydia p
8	132.5	4.6	1258	1	ICEN_ERWHE	P16239 erwinia her
9	131	4.6	2660	1	YEEJ_ECO57	Q8x8v7 escherichia
10	130.5	4.6	1034	1	ICEN_PANAN	Q47879 pantoea ana
11	130	4.6	1077	1	HLES_DROME	Q02308 drosophila
12	129.5	4.5	2411	1	DAB_DROME	P98081 drosophila
13	129	4.5	666	1	FLID_VIBCH	Q9kq63 vibrio chol
14	129	4.5	784	1	SP4_HUMAN	Q02446 homo sapien
15	129	4.5	1460	1	PMPC_CHLMU	Q9pjy1 chlamydia m
16	129	4.5	2090	1	HFC1_MESAU	P51611 mesocricetu
17	128.5	4.5	1902	1	P3P_LACLC	P15292 lactococcus
18	128	4.5	5703	1	MU5B_HUMAN	Q9hc84 homo sapien
19	127.5	4.5	436	1	Y868_CHLMU	Q9pjg1 chlamydia m
20	127.5	4.5	3726	1	TRX_DROME	P20659 drosophila
21	127	4.5	1337	1	DEXT_STRDO	P39653 streptococc
22	126	4.4	997	1	BIR1_SCHPO	O14064 schizosacch
23	125.5	4.4	1322	1	ICEA_PANAN	P20469 pantoea ana
24	125	4.4	1694	1	IGA0_HAEIN	P44969 haemophilus
25	125	4.4	1702	1	IGA2_HAEIN	P45384 haemophilus
26	125	4.4	1772	1	MSP1_PLAYO	P13828 plasmodium
27	125	4.4	1802	1	HKR1_YEAST	P41809 saccharomyc
28	125	4.4	1902	1	P1P_LACLC	P16271 lactococcus
29	124	4.3	1102	1	YG49_SCHPO	O60184 schizosacch
30	124	4.3	1902	1	P2P_LACLC	P15293 lactococcus
31	123.5	4.3	2483	1	PCX_DROME	P18490 drosophila
32	123	4.3	782	1	SP4_MOUSE	Q62445 mus musculu
33	123	4.3	1007	1	Y741_CHLMU	Q9pjt6 chlamydia m
34	123	4.3	1849	1	IGA4_HAEIN	P45386 haemophilus
35	122.5	4.3	2842	1	APC_RAT	P70478 rattus norv
36	122	4.3	2003	1	YDBA_ECOLI	P33666 escherichia
37	122	4.3	2035	1	HFC1_HUMAN	P51610 homo sapien
38	121.5	4.3	917	1	SMOO_HUMAN	P53814 homo sapien
39	121.5	4.3	1079	1	IF2P_SCHPO	Q10251 schizosacch
40	121.5	4.3	1189	1	YJH6_YEAST	P47035 saccharomyc
41	121	4.2	1150	1	APMU_PIG	P12021 sus scrofa
42	121	4.2	1306	1	MSB2_YEAST	P32334 saccharomyc
43	120.5	4.2	746	1	7UP2_DROME	P16376 drosophila
44	120	4.2	797	1	VGLX_HSVEB	P28968 equine herp
45	120	4.2	1508	1	BCSC_XANAC	P58938 xanthomonas

ALIGNMENTS

RESULT 1
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein homolog from lambdoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE000234; AAC74454.1; ALT_INIT.
DR EMBL; D90774; BAA14966.1; -.
DR EMBL; D90775; BAA14975.1; -.
DR PIR; G64887; G64887.
DR EcoGene; EG13370; stfR.
DR InterPro; IPR005003; Phage_fiber.
DR InterPro; IPR005068; Phage_fiber_2.
DR Pfam; PF03335; Phage_fiber; 6.
DR Pfam; PF03406; Phage_fiber_2; 1.
KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.

```
SQ   SEQUENCE   1120 AA;  113779 MW;  542E59D71EE795B4 CRC64;

  Query Match             5.1%;  Score 146.5;  DB 1;  Length 1120;
  Best Local Similarity   23.1%;  Pred. No. 0.46;
  Matches   92;  Conservative   52;  Mismatches  163;  Indels   91;  Gaps   17;

Qy         79 SEITLNDGFEVLH--------DHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQ------- 123
                | | | :||   |         |  |  |||  :|:      |::| :  |: :
Db         59 SVILLVEGFPPSHAGTITVYEDSQP-GTLNDFLGAM-----TEDDARPEALRRFELMVEE 112

Qy        124 --RNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILE 181
                ||    | |  :   |:  : |          |  |  | ||    |:|  |
Db        113 VARN---ASAVAQNTAAAKKSASDAS-------TSAREAATHA--ADAADSARA------ 154

Qy        182 LLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATG 241
                      |  ||::    :   |::|  :|  |: | | ::  |          : |
Db        155 --------ASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSK-------SAAAT 199

Qy        242 LIGLAATGIVQALA----LTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVN 297
                | | |   | |          : ||   :|| || : |     :||| :: :
Db        200 SAGAAKTSETNASASLQSAATSASTATTKASEAATSARDAAA----SKEAAKSSET---- 251

Qy        298 IDELGNAIPSGVLKDDVVANIEEQAKAAGEEAK-QQAIENNAQAQKKYDEQQAKRQEELK 356
                  ||  |         ::    | |||  ||  :  | ||:: :    | |      |
Db        252 -----NASSS-------ASSAASSATAAGNSAKAAKTSETNARSSETAAGQSASAAAGSK 299

Qy        357 VSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNT 416
                : |    | |    |    : |||     :  |   :: :| || :    |   |
Db        300 --TAAASSASAASTSAGQASASATAA----GKSAESAASSASTATTKAGEATEQASAAAR 353

Qy        417 PAQ----GNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
                |         :   || : || ::: ||::|:   ::|
Db        354 SASAAKTSETNAKASETSAESSKTAAASSASSAASSAS 391


RESULT 2
MUCS_BOVIN
ID   MUCS_BOVIN     STANDARD;      PRT;   563 AA.
AC   P98091;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Submaxillary mucin-like protein.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=90370871; PubMed=2204065;
RA   Bhargava A.K., Woitach J.T., Davidson E.A., Bhavanandan V.P.;
RT   "Cloning and cDNA sequence of a bovine submaxillary gland mucin-like
RT   protein containing two distinct domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6798-6802(1990).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC   -!- SIMILARITY: TO PORCINE APOMUCIN.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
CC   -!- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36192; AAA30657.1; -.
DR   PIR; A36054; A36054.
DR   InterPro; IPR006208; Cys_knot.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00007; Cys_knot; 1.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Repeat; Glycoprotein.
FT   DOMAIN       61    158       3 X 11 AA REPEATS.
FT   REPEAT       61     71       1.
FT   REPEAT      112    122       2.
FT   REPEAT      148    158       3.
FT   DOMAIN      338    404       VWFC.
FT   DOMAIN      471    555       CTCK.
FT   DISULFID    471    518       BY SIMILARITY.
FT   DISULFID    485    532       BY SIMILARITY.
FT   DISULFID    494    548       BY SIMILARITY.
FT   DISULFID    498    550       BY SIMILARITY.
FT   DISULFID     ?     554       BY SIMILARITY.
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    394    394       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   563 AA;  58913 MW;  AB326CD78E5FFFCF CRC64;

  Query Match             5.0%;  Score 142;  DB 1;  Length 563;
  Best Local Similarity   20.3%;  Pred. No. 0.35;
  Matches   75;  Conservative   48;  Mismatches  169;  Indels   78;  Gaps   13;

Qy         88 EVLHDHGPLDTLNRQIGSSVFRVET------------QEDGKHIAVGQRNGVETSVVLS 134
                :|| :: |   ::   |||     |              ||   :| || |  |
Db          2 KVLQENSPRHAIS---GSSHTEATTLIVSNSTSGTGLRPEDNTAVAGGQATGRVT----- 53

Qy        135 DQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKGTGESKGA 194
                        | | |  |     | :     : : |:| ||  :     || : |
Db         54 ------------GTTK-VIPGTTVAPGSSNTESTTSLGESRTRIGRI-----TGATTGT 94

Qy        195 GESKGVGELRESNSGA-ENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAATGIVQA 253
                :    |  :  |:||   ||    |::| :  |     |  |    | | :   |:
Db         95 SKRSSPGS-KTGNTGAISGTTVAPGSSNTGATTS-----LGSGETTQGGIKIVTMGVTTG 148

Qy        254 LALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKDD 313
                : |   :   | |    : ||  |  : |             : : |  |: | :|:
Db        149 TTIAPGSSNTKATTPTEVRTTTEVRTATETT--------TSRHSSDATGSGIQTGI---- 196

Qy        314 VVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGG 373
                        :|  :         |   |::      |  | :: ||   | ||   ::
Db        197 -------TGTGSGTTSSPGGFNAEATTFKEH-----VRTTETRILSGTTRGRSGTTVIPE 244

Qy        374 GIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMES 433
                      :  : |       ||:|   :   |   :  : |  :  |      ||   | :
Db        245 SSNTGTSTGVGR------QTSTAVVSGRVTGV-SESSSPGTSKEASETTTGPGISTTGST 297

Qy        434 RRSSMASTSS 443
                :|:  :|||
Db        298 SKSNRITTSS 307


RESULT 3
STAU_DROME
ID   STAU_DROME     STANDARD;      PRT;  1026 AA.
AC   P25159;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Maternal effect protein staufen.
GN   STAU.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91300552; PubMed=1712672;
RA   St Johnston D., Beuchle D., Nuesslein-Volhard C.;
RT   "Staufen, a gene required to localize maternal RNAs in the Drosophila
RT   egg.";
RL   Cell 66:51-63(1991).
RN   [2]
RP   STRUCTURE BY NMR OF 579-646.
RX   MEDLINE=95354674; PubMed=7628456;
RA   Bycroft M., Grunert S., Murzin A.G., Proctor M., St Johnston D.;
RT   "NMR solution structure of a dsRNA binding domain from Drosophila
RT   staufen protein reveals homology to the N-terminal domain of
RT   ribosomal protein S5.";
RL   EMBO J. 14:3563-3571(1995).
RN   [3]
RP   ERRATUM.
RA   Bycroft M., Grunert S., Murzin A.G., Proctor M., St Johnston D.;
RL   EMBO J. 14:4385-4385(1995).
RN   [4]
RP   CHARACTERIZATION OF DRBM DOMAINS.
RX   MEDLINE=20183617; PubMed=10716936;
RA   Micklem D.R., Adams J., Grunert S., St Johnston D.;
RT   "Distinct roles of two conserved Staufen domains in oskar mRNA
RT   localization and translation.";
RL   EMBO J. 19:1366-1377(2000).
CC   -!- FUNCTION: REQUIRED BOTH FOR THE LOCALIZATION OF MATERNAL
CC       DETERMINANTS TO THE POSTERIOR POLE OF THE DROSOPHILA EGG AND
CC       FOR BICOID RNA TO LOCALIZE CORRECTLY TO THE ANTERIOR POLE.
CC       OSKAR PROTEIN IS REQUIRED TO KEEP OSKAR RNA AND STAUFEN PROTEIN AT
CC       THE POSTERIOR POLE.
CC   -!- TISSUE SPECIFICITY: POLAR GRANULES AT THE POSTERIOR POLE OF THE
CC       OOCYTE, AND BY THE TIME THE EGG IS LAID, AT THE ANTERIOR POLE.
CC   -!- DOMAIN: CONTAINS A PROLINE-RICH DOMAIN. THE INSERTION OF THIS
CC       DOMAIN IN THE DRBM 2 DOMAIN IS REQUIRED FOR STAU-OSKAR MRNA
CC       LOCALIZATION.
CC   -!- SIMILARITY: Contains 5 DRBM (double-stranded RNA-binding) domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69111; AAA73062.1; -.
DR   PIR; A40315; A40315.
DR   PDB; 1STU; 31-JUL-95.
DR   PDB; 1EKZ; 21-AUG-00.
DR   FlyBase; FBgn0003520; stau.
DR   GO; GO:0045179; C:apical cortex; IDA.
DR   GO; GO:0003730; F:mRNA 3' UTR binding activity; IDA.
DR   GO; GO:0045034; P:neuroblast cell division; NAS.
DR   GO; GO:0045450; P:pole plasm bicoid mRNA localization; NAS.
DR   GO; GO:0007316; P:pole plasm RNA localization; IMP.
DR   InterPro; IPR001159; DS_RBD.
DR   Pfam; PF00035; dsrm; 5.
DR   SMART; SM00358; DSRM; 4.
DR   PROSITE; PS50137; DS_RBD; 5.
KW   Developmental protein; RNA-binding; Repeat; 3D-structure.
FT   DOMAIN      311    378       DRBM 1.
FT   DOMAIN      398    557       DRBM 2 (ATYPICAL).
FT   DOMAIN      578    645       DRBM 3.
FT   DOMAIN      711    781       DRBM 4.
FT   DOMAIN      951   1018       DRBM 5.
FT   DOMAIN        3     77       GLN/HIS/PRO-RICH.
FT   DOMAIN       42     46       POLY-GLN.
FT   DOMAIN       47     51       POLY-PRO.
FT   DOMAIN       71     77       POLY-GLN.
FT   DOMAIN      451    455       POLY-PRO.
FT   HELIX       580    590
FT   TURN        591    591
FT   STRAND      595    598
FT   STRAND      609    616
FT   TURN        617    618
FT   STRAND      619    627
FT   TURN        628    629
FT   HELIX       630    644
SQ   SEQUENCE   1026 AA;  110347 MW;  AE5B97624BBF7D0B CRC64;

  Query Match             4.9%;  Score 140.5;  DB 1;  Length 1026;
  Best Local Similarity   20.0%;  Pred. No. 0.92;
  Matches   97;  Conservative   74;  Mismatches  207;  Indels  107;  Gaps   18;

Qy          2 PIG-NLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPL-------GSRALFT---- 49
              |:| :: | |  |   |  |  ||:     |: : :     |         :|||
Db        501 PLGAHVHHGP--NGPFPSVPTPPSKITLFVGKQKFVGIGRTLQQAKHDAAARALQVLKTQ 558

Qy         50 -------PVRNSMADSGDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQ 102
                    : :|| | || ::      |:: :     :  :   |:|| : ||     |
Db        559 AISASEEALEDSM-DEGDKKS------PISQVHEIGIKRNMTVHFKVLREEGPAHMKNFI 611

Qy        103 IGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKF----VFTGGR 158
                  |  : |: :|    | ::   |  :|    |  :| : |  :        | | |:
Db        612 TACIVGSIVTEGEGNGKKVSKKRAAEKMLV----ELQKLPPLTPTKQTPLKRIKVKTPGK 667

Qy        159 GGAGHAMVTVAS------------------------DITEARQRILELLEPKGTGESKG 193
                ||     :| |                      |: :|   | :|:: : | :
Db        668 SGAAAREGSVVSGTDGTMQTGKPERRKRLNPPKDKLIDMDDADNPITKLIQLQQTRK--- 724

Qy        194 AGESKGVGELRESNSG-----AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAAT 248
                | : : ||   |         |   |  | ||:   : |  ||    |  | |
Db        725 --EKEPIFELIAKNGNETARRREFVMEVSASGSTARGTGNSKK-LAKRNAAQALFEL--- 778

Qy        249 GIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSG 308
              ::|: :||      | :  ::    :||  :|  | :      || :       |
Db        779 --LEAVQVTP------TNETQSSEECCTSATMSAVTAPAVEATAEGKVPMVATPVGPMPG 830

Qy        309 VL-----------KDDVV---ANIE--------EQAKAAGEEAKQQAIENNAQAQKKYDE 346
              :|           :| :|   :|:|        | | || | :   |   ::      |
Db        831 ILILRQNKKPAKKRDQIVIVKSNVESKEEEANKEVAVAAEENSNNSANSGDSSNSSSGDS 890

Qy        347 QQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSAR 406
              |  :   | :::  |   ||       :|       | :::    | ::: :|: | :
Db        891 QATEAASESALNTSTGSNTSGVSSNSSNVGANTDGNNHAESK--NNTESSSNSTSNTQSA 948

Qy        407 TVENK 411
                |||
Db        949 GVHMK 953


RESULT 4
US45_LACLC
ID   US45_LACLC     STANDARD;      PRT;   461 AA.
AC   P22865;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Secreted 45 kDa protein precursor.
GN   USP45.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1359;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RC   STRAIN=MG1363;
RX   MEDLINE=91071599; PubMed=2123812;
RA   van Asseldonk M., Rutten G., Oteman M., Siezen R.J., de Vos W.M.,
RA   Simons G.;
RT   "Cloning of usp45, a gene encoding a secreted protein from
RT   Lactococcus lactis subsp. lactis MG1363.";
RL   Gene 95:155-160(1990).
CC   -!- SIMILARITY: TO E.FAECIUM P54.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60178; AAA25230.1; -.
DR   EMBL; A17083; CAA01320.1; -.
DR   Pfam; PF05257; AXE; 1.
KW   Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    461       SECRETED 45 kDa PROTEIN.
FT   DOMAIN      299    314       POLY-SER.
SQ   SEQUENCE   461 AA;  47387 MW;  51493C42224F3C03 CRC64;

  Query Match             4.8%;  Score 136.5;  DB 1;  Length 461;
  Best Local Similarity   19.1%;  Pred. No. 0.56;
  Matches   84;  Conservative   69;  Mismatches  163;  Indels  123;  Gaps   14;

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              :: :   | |  : :::   ::    |  ::  : ||||     : : :|       :||
Db         32 DIAKQDATISSAQSAKAQAQAQVDSLQSKVDSLQQKQTSTKAQIAKIESELKALNAQIAT 91

Qy        241 GLIGL-AATGIVQALALTPEPDSPTTTDPDA-------------AASATETATRDQL-- 283
                :   |  ::| | : : :|  |   ||               | || ::   |:
Db         92 LNESIKERTKTLEAQARSAQVNSSATNYMDAVVNSKSLTDVIQKVTAIATVSSANKQILE 151

Qy        284 -----TKEAFQNPDNQKVNIDE---LGNAIPS---------GVLKDDVVANIEEQAKAAG 326
                  ||  |  :  | | ::   |  :: |           ||   || :  ||  |
Db        152 QQEKEQKELSQKSETVKKNYNQFVSLSQSLDSQAQELTSQQAELK---VATLNYQATIAT 208

Qy        327 EEAKQQAIENNAQAQKKYDEQQAKRQE--ELKVSSGAGYGLSGALILGGGIGVAVTAALH 384
              : |:||: :   | :|  :: ||:|   | :    |    :         :  | ::|
Db        209 AQDKKQALLDEKAAAEKAAQEAAKKQAAYEAQQKEAAQAQAASTAATAKAVEAATSSASA 268

Qy        385 RKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSST 444
              :|  : :|:|  ||:  ||     |  :|:: :  :     |  :  |  |:    :::
Db        269 SSSQAPQVSTSTDNTTSNASASNSSNSSSNSSSSSSS----SSSSSSSSSNSNAGGNTNS 324

Qy        445 FFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDN 504
              |  : ||                                                ||
Db        325 GTSTGNTGG-------------------------------------------------TTTG 337

Qy        505 GARL----LGNPSAG------------IQSTYARLALSG------------GLRHDMG-- 534
              |: :    :||| ||             |  | |  : |            |: | :|
Db        338 GSGINSSPIGNPYAGGGCTDYVWQYFAAQGIYIRNIMPGNGGQWASNGPAQGVLHVVGAA 397

Qy        535 -GLTGGSNSA--VNTSNNP 550
               |:   | ||  |  :|:|
Db        398 PGVIASSFSADFVGYANSP 416


RESULT 5
YS89_CAEEL
ID   YS89_CAEEL     STANDARD;      PRT;  3178 AA.
AC   Q09624; Q09625; Q969D4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein ZK945.9 in chromosome II.
GN   ZK945.9/ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   -!- SIMILARITY: Contains 1 PLAT domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z48544; CAB70192.1; -.
DR   EMBL; Z48582; CAB70192.1; JOINED.
DR   EMBL; Z48582; CAB70201.1; -.
DR   EMBL; Z48544; CAB70201.1; JOINED.
DR   WormPep; ZK945.9; CE25697.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR001024; Lipoxygenase_LH2.
DR   InterPro; IPR003915; PKD_2.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF01477; PLAT; 1.
DR   PRINTS; PR01433; POLYCYSTIN2.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00308; LH2; 1.
DR   PROSITE; PS50095; PLAT; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     13     30       POTENTIAL.
FT   TRANSMEM     51     73       POTENTIAL.
FT   TRANSMEM   2139   2161       POTENTIAL.
FT   TRANSMEM   2348   2367       POTENTIAL.
FT   TRANSMEM   2390   2412       POTENTIAL.
FT   TRANSMEM   2451   2468       POTENTIAL.
FT   TRANSMEM   2483   2505       POTENTIAL.
FT   TRANSMEM   2567   2589       POTENTIAL.
FT   TRANSMEM   2836   2858       POTENTIAL.
FT   TRANSMEM   2939   2961       POTENTIAL.
FT   TRANSMEM   2976   2998       POTENTIAL.
FT   TRANSMEM   3038   3060       POTENTIAL.
FT   DOMAIN      266   1196       SER/THR-RICH.
FT   DOMAIN     1105   1241       GLY/SER-RICH.
FT   DOMAIN     2071   2120       GPS.
FT   DOMAIN     2182   2305       PLAT.
SQ   SEQUENCE   3178 AA;  344726 MW;  F8239436D03666CD CRC64;

  Query Match             4.7%;  Score 133.5;  DB 1;  Length 3178;
  Best Local Similarity   19.7%;  Pred. No. 9.8;
  Matches   82;  Conservative   76;  Mismatches  176;  Indels   83;  Gaps   14;

Qy         68 GLPVNPMRLAASEITLND--------GFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHI 119
              || :|   :  :|:  :|         :|| :| |    ::|  ||  :    | |  :
Db        137 GLFLNSTWITLNEVNDDDEISIAVEAKYEVCYDDG----IDRCDGSLWW---LQVGGNEM 189

Qy        120 A-VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQR 178
              | :| |   |:  :  ::||||      |   :|                   :: |::::
Db        190 ALLGYREKCESGEI--NEEYARRMCKRPYRSEK----------------STAISDSQGV 230

Qy        179 ILELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTV 238
                :    ||    : :  : |   ||      | : |   |  |||:  : |       ||
Db        231 YYDGQVLKGVRAKQFSMRTSGSPTLRRMKRDAGDNTCDYTIESTSTSTTTPTTTTVTSTV 290

Qy        239 -ATGLIGLAATGIVQALALTPEPDSPTTT--------DPDAAASATETATRDQLTKEAFQ 289
               :|  :  : : :  |:: :    | :||           |: | : |:|  | :
Db        291 TSTTTVPTSTSTVTTAMSTSTSTPSTSTTIESTSTTFTSTASTSTSSTSTTQQSSSTITS 350

Qy        290 NPDNQKVNIDELGNAIPSGVLKD--DVVANIEEQAKAAG-EEAKQQAIENNAQAQKKYDE 346
              :| :       |  :||:    :    :::: : |  :  :|                |
Db        351 SPSS-----TTLSTSIPTTTTPEITSTLSSLPDNAICSYLDETTTSTTFTTTMLTSTTTE 405
```

```
Qy        347 QQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTT--S 404
              : :      :|:|                     |::     :|    ||:| :|:||  |
Db        406 EPSTSTTTTEVTS-------------------TSSTVTTTEPTTTLTTSTASTSTTEPS 445

Qy        405 ARTVENKPA---------NNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIG 452
              ||   |:         ::: :   | || | ::  :  ||  :||:|   ||: |
Db        446 TSTVTTSPSTSPVTSTVTSSSSSSTTVTTPTSTESTSTSPSSTVTTSTTAPSTSTTG 502


RESULT 6
Y579_CHLTR
ID   Y579_CHLTR     STANDARD;      PRT;   439 AA.
AC   O84583;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein CT579.
GN   CT579.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/Cx;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CT579/TC0868
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001328; AAC68181.1; -.
DR   PIR; E71497; E71497.
DR   PHCI-2DPAGE; O84583; -.
KW   Complete proteome.
SQ   SEQUENCE   439 AA;  44023 MW;  F078A4149E908B68 CRC64;

  Query Match             4.7%;  Score 133;  DB 1;  Length 439;
  Best Local Similarity   23.8%;  Pred. No. 0.84;
  Matches   73;  Conservative   40;  Mismatches  100;  Indels   94;  Gaps   14;

Qy        182 LLEPKGTGESKGAGESK----GVGELRESNSGAENTTE----TQTSTS-------TSSLR 226
              | : :|||  || |  ::    |  |  :::||   ||     :||:|:         |
Db         53 LQDAQGTGGSKKATANQTSKSGKSEKAQASSGTSTTTSVAQASQTATAQAVHGARDSGFN 112

Qy        227 SDPKLWL--ALGTVATGLI------GLAATGIVQALALTPEPDSPTTTDPDAAASATETA 278
              ||    |   ||   |::       ||  |::  |               | |||   :
Db        113 SDGSATLPSPTGTEVNGVVLRKGMGTLALMGLIMTLL-------------AQASAKSWS 158

Qy        279 TRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNA 338
              :  |   :| |   ||      |:||| :              ||:|   :|:|  |
Db        159 SSFQQQNQAIQ---NQVAMAPEIGNAIRT---------QANHQAQATELQAQQSLISG-- 204

Qy        339 QAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGI-------GVAVTAALHRKNQPVE 391
                            :::  |: :|    :||||       |   :||
Db        205 ------------------ITNIVGFAVS----VGGGILSASKSLGGLKSAAF------TN 236

Qy        392 QTTTTTTTTTTTSARTVE---NKPANNTPAQGNVDTPGSEDTMESRRSSMA-----STSS 443
              :| : ||: |:: |:|     :  |    | |   | |:      |  : :       | |
Db        237 ETASATTSATSSLAKTATSALDDVAGTATAVGAKATSGAASAASSAATKLTQNMAESASK 296

Qy        444 TFFDTSS 450
              |   |:|
Db        297 TLSQTAS 303


RESULT 7
Y572_CHLPN
ID   Y572_CHLPN     STANDARD;      PRT;   755 AA.
AC   Q9Z7Y1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein CPn0572/CP0177/CPj0572.
GN   CPN0572 OR CP0177 OR CPJ0572.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0572/CT456/TC0741
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001642; AAD18712.1; -.
DR   EMBL; AE002178; AAF38051.1; -.
DR   EMBL; AP002547; BAA98778.1; -.
DR   PIR; B72061; B72061.
DR   PIR; H86561; H86561.
DR   PHCI-2DPAGE; Q9Z7Y1; -.
DR   TIGR; CP0177; -.
KW   Complete proteome.
SQ   SEQUENCE   755 AA;  77578 MW;  E57D165BE973AC81 CRC64;

  Query Match             4.6%;  Score 132.5;  DB 1;  Length 755;
  Best Local Similarity   21.1%;  Pred. No. 1.8;
  Matches  104;  Conservative   57;  Mismatches  166;  Indels  167;  Gaps   23;

Qy          5 NLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGP----------------------L 42
              ::|    :|    |||      :  ||     : |||                       |
```

```
Db        325 DVGGLKGINWKTTPAPDFSFINETPGGGAHSTSHTGPGTPVGATVVPNVNVNLGGIKVDL 384

Qy         43 GSRAL-------------FTPVRNSMADSGDNRAS-----------------DVPGL--- 69

Db        385 GGINLGGITTNVTTEEGGGTNITSTKSTSTDDKVSITSTGSQSTIEEDTIQFDDPGQGED 444

Qy         70 ------------PVNPMRLAASEI------TLNDGFEVLHDHGPLDTLNRQIGSSVFRVE 111

Db        445 DNAIPGTNTPPPPGPPPNLSSSRLLTISNASLNQVLQNVRQH--LNTAYDSNGNSV--SD 500

Qy        112 TQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGG-RGGAGHAMVTVAS 170

Db        501 LNQDLGQVVKNSENGVNFPTVILPKTTG---DTDPSGQ----ATGGVTEGGGH------- 546

Qy        171 DITEARQRILELLEPKGTGESKGAGE------SKGVGELRESNSGAEN----------T 213

Db        547 -IRNIIQR-----NTQSTGQSEGATPTPQPTIAKIVTSLRKANVSSSSVLPQPQVATTIT 600

Qy        214 TETQT-STSTSSLRSDPKLWLALGT-VATGLIGLAATGIVQALALTPEPDSPTTTDPDAA 271

Db        601 PQARTASTSTTSIGTGTESTSTTSTGTGTGSVSTQSTGV----------GTPTTTTRSTG 650

Qy        272 ASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQ 331

Db        651 TSATTTT-----SSASTQTPQ----------APLPSGTRH---VATISLVRNAAG----- 687

Qy        332 QAIENNAQAQKKYDEQQAKRQEELKV-SSGAGYGLSGALI------LGGGIGVAVTAALH 384

Db        688 ----------RSIVLQQGGRSQSFPIPPSGTGTQNMGAQLWAAASQVASTLGQVVNQAAT 737

Qy        385 RKNQPVEQTTTTTT 398

Db        738 AGSQPSSRRSSPTS 751
```

```
RESULT 8
ICEN_ERWHE
ID   ICEN_ERWHE     STANDARD;      PRT;  1258 AA.
AC   P16239;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   ICEE.
OS   Erwinia herbicola.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Pantoea.
OX   NCBI_TaxID=549;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M1;
RX   MEDLINE=90152370; PubMed=2515997;
RA   Warren G.J., Corotto L.V.;
RT   "The consensus sequence of ice nucleation proteins from Erwinia
RT   herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL   Gene 85:239-242(1989).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC       A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC       PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M26382; AAA24823.1; -.
DR   PIR; JQ0188; JQ0188.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 65.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 45.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      162   1217       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1258 AA;  125084 MW;  590E8A130077FBD4 CRC64;

  Query Match             4.6%;  Score 132.5;  DB 1;  Length 1258;
  Best Local Similarity   20.4%;  Pred. No. 3.4;
  Matches  110;  Conservative   62;  Mismatches  226;  Indels  141;  Gaps   20;

Qy         26 TDGAGGRGQLI---NSTGPLGSRALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEI- 81

Db        183 TETAGDSSTLIAGYGSTGTAGADSTLVAGYGSTQTAGE-ESSQMAGYGSTQTGMKGSDLT 241

Qy         82 -------TLNDGFEVLHDHGPLDT------LNRQIGSSVFRVETQEDGKHIAVGQRN--- 125

Db        242 AGYGSTGTAGDDSSLIAGYGSTQTAGEDSSLTAGYGST----QTAQKGSDLTAGYGSTGT 297

Qy        126 -GVETSVVLSDQEYARLQSIDPE--------------GKDKFVFTGGRGGAG------- 162

Db        298 AGADSSLIAG---YGSTQTAGEESTQTAGYGSTQTAQKGSDLTAGYGSTGTAGDDSSLIA 354

Qy        163 ----------HAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGAEN 212

Db        355 GYGSTQTAGEDSSLTAGYGSTQTAQKGSDLTAGYGSTGTAGADSSLIAGYGSTQTAGEES 414

Qy        213 T-TETQTSTSTSSLRSDPKLWLALGTVATGLIG-----LAATGIVQALALTPEPDSPTT- 265

Db        415 TQTAGYGSTQTAQKGSD----LTAGYGSTGTAGDDSSLIAGYGSTQ----TAGEDSSLTA 466

Qy        266 -----------TDPDAAASATETATRDQLTKEAF------------------QNPDNQK 295

Db        467 GYGSTQTAQKGSDLTAGYGSTSTAGYESSLIAGYGSTQTAGYGSTLTAGYGSTQTAQNES 526

Qy        296 VNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEEL 355

Db        527 DLITGYGSTSTAGA----------NSSLIAGYGSTQTASYNSVLTAGYGSTQTAREGSDL 576

Qy        356 KVSSGAGYGLSGAL----ILGGGIGVAVTAALH-----------RKNQPVEQTTTTTTT 399

Db        577 T----AGYGSTGTAGSDSSIIAGYGSTQTASYHSSLTAGYGSTQTAREQSVLTTGYGSTS 632

Qy        400 TTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSS-----MASTSSTFFDTSSIGG 453

Db        633 TAGADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQEGSDLTAGYGSTSTAGADSSLIAG 691
```

```
RESULT 9
YEEJ_ECO57
ID   YEEJ_ECO57     STANDARD;      PRT;  2660 AA.
AC   Q8X8V7; Q8X2B9; Q8X2C0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yeeJ.
GN   Z3135 OR ECS2775/ECS2776.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
```

```
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: Contains 16 Big-1 domains.
CC   -!- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1315.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005423; AAG57041.1; -.
DR   EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR   EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR   InterPro; IPR003344; Big_1.
DR   InterPro; IPR003535; Intimin.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF02369; Big_1; 16.
DR   PRINTS; PR01369; INTIMIN.
DR   SMART; SM00634; BID_1; 16.
DR   SMART; SM00089; PKD; 8.
KW   Hypothetical protein; Repeat; Complete proteome.
FT   DOMAIN      738    834       BIG-1 1.
FT   DOMAIN      840    929       BIG-1 2.
FT   DOMAIN      931   1033       BIG-1 3.
FT   DOMAIN     1042   1132       BIG-1 4.
FT   DOMAIN     1134   1236       BIG-1 5.
FT   DOMAIN     1245   1335       BIG-1 6.
FT   DOMAIN     1337   1439       BIG-1 7.
FT   DOMAIN     1448   1539       BIG-1 8.
FT   DOMAIN     1548   1652       BIG-1 9.
FT   DOMAIN     1653   1750       BIG-1 10.
FT   DOMAIN     1751   1855       BIG-1 11.
FT   DOMAIN     1856   1957       BIG-1 12.
FT   DOMAIN     1963   2056       BIG-1 13.
FT   DOMAIN     2065   2156       BIG-1 14.
FT   DOMAIN     2157   2252       BIG-1 15.
FT   DOMAIN     2254   2355       BIG-1 16.
SQ   SEQUENCE   2660 AA;  280062 MW;  01EB92A08F5C09D2 CRC64;

  Query Match             4.6%;  Score 131;  DB 1;  Length 2660;
  Best Local Similarity   19.7%;  Pred. No. 11;
  Matches  123;  Conservative   88;  Mismatches  224;  Indels  190;  Gaps   25;

Qy         33 GQLINSTGPLGSRALFTPVRNSMADS----------------------GDNRASDVPG 68
              |  :| | |    | ||   | :| :                        |:| |||
Db       1176 GITVNFTMPQDVAANFTLENNGIAITQANGEAHVTLKGKKAGTHTVTATLGNNNASDAQP 1235

Qy         69 L--------PVNPMRLAASEI------------TLNDGF-------EVLHDHGPLDTLNR 101
              :         |  :: : :||            |: | |       :|     | ||
Db       1236 VTFVADKDSAVVVLQTSKAEIIGNGVDETTLTATVKDPFDNAVKDLQVTFSTNPADT--- 1292

Qy        102 QIGSSVFRVETQEDG--KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRG 159
              |:  |  :  | : |  :    |  || |:           ::  | | :

Db       1293 QLSQS--KSNTNDSGVAEVTFKGTVLGVHTA----------EATLPNGNNDTKIVNIAP 1339

Qy        160 GAGHAMVT-------VASDITEARQRILELLEPKG-------------------TGESK 192
              | :| ||        | :: ::: |     : :|                       | |:
Db       1340 DASNAQVTLNIPAQQVVTNNSDSVQLTATVKDPSNHPVAGITVNFTMPQDVAANFTLENN 1399

Qy        193 GAGESKGVGELRESNSGAENTTETQTSTSTSSLRSD--PKLWLALGTVATGLIGLAATGI 250
              |   ::  ||   :  | :  | | |:| :::  ||  |  ::|  | |  :: ::   |
Db       1400 GIAITQANGEAHVTLKGKKAGTHTVTATLSNNNTSDSQPVTFVADKTSALVVLQISKNEI 1459

Qy        251 VQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVL 310
                              :  ||| |||     |: |   ||:  |:    :   ||:
Db       1460 T----------------GNGVDSATLTAT----VKDQF---DNEVNNLPVTFSTASSGLT 1496

Qy        311 KDDVVANIEE----QAKAA----GEEAKQQAIENNAQAQKKY-----DEQQAKRQEELKV 357
                :|  |    ||  |    ||:    :: ||  :  |       |  ||  |   |
Db       1497 LTPGESNTNESGIAQATLAGVAFGEQTVTASLANNGASDNKTVHFIGDTAAAKIIELTPV 1556

Qy        358 SSGAGYGLSGALILG---GGIGVAVTAALHRKNQ-PVEQTTTTTTTTTTTSARTVENKPA 413
                      ::| |       |  :|| :   |  ||:  |   |:   |:  |   :
Db       1557 PD--------SIIAGTPQNSSGSVITATVVDNNGFPVKGVTVNFTSNAATAEMTNGGQAV 1608

Qy        414 NN---------TPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKH 464
                |          |  : ::::    ||:|:   : :|| ||  : ::      : :|
Db       1609 TNEQGKATVTYTNTRSSIESGARPDTVEASLENGSSTLSTSINVNADASTAHLTLL---- 1664

Qy        465 RCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA 524
                                      :: |:     |||:    :  |   |:      |:
Db       1665 ----------------------QALFDTVS--AGDTTNLYIEVKDNYGNGVPQQEVTLS 1699

Qy        525 LSGGLRHDMGGLTGGSNSAVNTSNN 549
              :|        |:|  ||:|: |:|:
Db       1700 VS-----PSEGVT-PSNNAIYTTNH 1718


RESULT 10
ICEN_PANAN
ID   ICEN_PANAN     STANDARD;      PRT;  1034 AA.
AC   Q47879;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein inaU.
GN   INAU.
OS   Pantoea ananas (Erwinia uredovora).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Pantoea.
OX   NCBI_TaxID=553;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KUIN-3;
RX   MEDLINE=94264407; PubMed=7764866;
RA   Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT   "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT   uredovora.";
RL   Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC       A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC       PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D14992; BAA03636.1; -.
DR   PIR; JC2143; JC2143.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 51.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 34.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      162    993       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1034 AA;  103378 MW;  FA222523D333EADD CRC64;

  Query Match              4.6%;  Score 130.5;  DB 1;  Length 1034;
  Best Local Similarity   20.9%;  Pred. No. 3.5;
  Matches  114;  Conservative   64;  Mismatches  229;  Indels  139;  Gaps   23;

Qy         26 TDGAGGRGQLI---NSTGPLGSRALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEI- 81
              |: ||    ||    |||  || :       |   :|:  :| : |       :  |::
Db        183 TETAGDSSTLIAGYGSTGTAGSDSTLVAGYGSTQTAGE-ESSQMAGYGSTQTGMKGSDLT 241

Qy         82 -------TLNDGFEVLHDHGPLDT------LNRQIGSSVFRVETQEDGKHIAVGQRN--- 125
                    |  |   ::  :|   |       |     ||:    :| : |  :  |   :
Db        242 AGYGSTGTAGDDSSLIAGYGSTQTAGEDSSLTAGYGST----QTAQKGSDLTAGYGSTGT 297

Qy        126 -GVETSVVLSDQEYARLQSIDPE--------------GKDKFVFTGGRGGAG------- 162
               | ::|::      |   |:  |                 | |     |  | ||
Db        298 AGADSSLIAG---YGSTQTAGEESTQTAGYGSTQTAQKGSDLTAGYGSTGTAGDDSSLIA 354

Qy        163 ----------HAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGAEN 212
                       : :|     |:  |:  :|    |:  : || |   |     :| |:
Db        355 GYGSTQTAGEDSSLTAGYGSTQTAQKGSDLTAGYGSTGTAGADSSLIAGYGSTQTAGEES 414

Qy        213 T-TETQTSTSTSSLRSDPKLWLALGTVATGLIG-----LAATGIVQALALTPEPDSPTT- 265
              | |    || |:   ||     |  |  :||  |     :|  | |    |   ||  |
Db        415 TQTAGYGSTQTAQKGSD----LTAGYGSTGTAGDDSSLIAGYGSTQ----TAGEDSSLTA 466

Qy        266 -----------TDPDAAASATETATRDQLTKEAF------------------QNPDNQK 295
                        :|  |   :| ||  :     :                      |   |:
Db        467 GYGSTQTAQKGSDLTAGYGSTSTAGYESSLIAGYGSTQTAGYGSTLTAGYGSTQTAQNES 526

Qy        296 VNIDELGNAIPSGVLKDDVVANIEEQAKA-------AGEEAKQQAIENNAQAQKKYDEQQ 348
               |    |:   :|      ::|       |          ||  : | | | :       |
Db        527 DLITGYGSTSTAGA-NSSLIAGYGSTQTASYNSVLTAGYGSTQTAREGSDLTAGYGSTQT 585

Qy        349 AKRQEELKVSSG----AGYGLSGALILGGGIGVAVTAALHR-------KNQPVEQTTTTT 397
              |:   :|    |     |||  |     |  | |  ||  |                |  :: :  |
Db        586 AQENSDLTTGYGSTSTAGYDSS----LIAGYGSTQTAGYHSILTAGYGSTQTAQERSDLT 641

Qy        398 T---TTTTTSART--VENKPANNTPAQGNVDTPGSEDTMESRRSS-----MASTSSTFFD 447
              |   :|:|  | :  :    :  |     :: | |   |  :: :|       |||:  ::
Db        642 TGYGSTSTAGADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQENSDLTTGYGSTSTAGYE 701

Qy        448 TSSIGG 453
               :| | |
Db        702 SSLIAG 707


RESULT 11
HLES_DROME
ID   HLES_DROME     STANDARD;      PRT;  1077 AA.
AC   Q02308; Q9VDK0; Q9VDK1;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hairless protein.
GN   H OR CG5460.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=92387549; PubMed=1516831;
RA   Bang A.G., Posakony J.W.;
RT   "The Drosophila gene Hairless encodes a novel basic protein that
RT   controls alternative cell fates in adult sensory organ development.";
RL   Genes Dev. 6:1752-1769(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=93041287; PubMed=1419850;
RA   Maier D., Stumm G., Kuhn K., Preiss A.;
RT   "Hairless, a Drosophila gene involved in neural development, encodes
RT   a novel, serine rich protein.";
RL   Mech. Dev. 38:143-156(1992).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
CC       DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
CC       FATES BY THE TRICHOGEN (SHAFT) / TORMOGEN (SOCKET) SISTER CELL
CC       PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
CC       OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
CC       TORMOGEN FATE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q02308-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q02308-2; Sequence=VSP_006952;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
CC       DISCS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M95192; AAA28607.1; ALT_INIT.
DR   EMBL; X67239; CAA47664.1; -.
DR   EMBL; AE003731; AAF55790.1; -.
DR   EMBL; AE003731; AAF55791.1; -.
DR   PIR; A44067; A44067.
DR   FlyBase; FBgn0001169; H.
DR   GO; GO:0003714; F:transcription co-repressor activity; IDA.
DR   GO; GO:0007219; P:N receptor signaling pathway; NAS.
DR   GO; GO:0008052; P:sensory organ determination; IMP.
KW   Developmental protein; Nuclear protein; DNA-binding;
KW   Alternative splicing.
FT   DOMAIN      115    123       THR-RICH.
FT   DOMAIN      642    648       POLY-SER.
FT   DOMAIN      879    891       POLY-ALA.
FT   DOMAIN      937    946       POLY-ALA.
FT   DOMAIN      964    974       ALA-RICH.
FT   DOMAIN      979   1008       HIS/PRO-RICH (PRD MOTIF).
FT   VARSPLIC      1     18       Missing (in isoform 2).
FT                                /FTId=VSP_006952.
FT   CONFLICT    151    151       S -> A (IN REF. 2).
FT   CONFLICT    680    680       S -> F (IN REF. 3).
FT   CONFLICT    702    703       QH -> LL (IN REF. 2).
FT   CONFLICT    891    891       A -> R (IN REF. 2).
FT   CONFLICT    964    967       AAVA -> RLLP (IN REF. 2).
FT   CONFLICT    974    974       MISSING (IN REF. 2).
SQ   SEQUENCE   1077 AA;  111039 MW;  A94BF1A27579E2F1 CRC64;

  Query Match             4.6%;  Score 130;  DB 1;  Length 1077;
  Best Local Similarity   18.9%;  Pred. No. 3.9;
  Matches  145;  Conservative   66;  Mismatches  244;  Indels  312;  Gaps   29;

Qy          7 GHNPNVNNSIPPAPP---------------LPSQTDGAGGRGQLINSTGPLGS--RALF 48
              | | : : | || ||               || :     | | ::  |  |   |
Db         92 GLNGSSSTSYPPLPPPLPANLSRTTTPTTTTPSSSSSTASNGFLPHAKTPKSSSIMAAS 151

Qy         49 TPVRNSMADSGDNRAS-DVPG--------------LPVNPMRLAASEIT-----LNDGF 87
                |  |:  :  :: : || |               ||   :  ||:         | |
Db        152 AAVAASVVGATASKPTIDVLGGVLDYSSLGGAATGSLPTTAVVAAAAGTAKIGKGSNSGG 211

Qy         88 EVLHDHGPLDTL-NRQIGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDP 146
                    |: |  |   |    |::  :||| |                | || :  |
Db        212 SFDMGRTPISTHGNNSWGGYGGRLQFFKDGKFIL---------------ELARSKDGDK 255

Qy        147 EGKDKFVFTGGR--GGAGHAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELR 204
                |        |    |  | ||  | :| |         ||
Db        256 SGWVSVTRKTFRPPSAATSATVTPTSAVTTA--------YPKN---------------- 290

Qy        205 ESNSGAENTTETQTSTSTSSLRSDP----------------KLWLALGTVATGLIGLA 246
                    ||:|    |   ||::| |                :| |         :| |
Db        291 ------ENSTSLSFSDDNSSIQSSPWQRDQPWKQSRPRRGISKELSLFFHRPRNSTLGRA 344

Qy        247 ATGIVQALALTPEPDSPTTTDPD-------------------AAASATE---TATRDQL 283
              |  :  |      |  | || ^|                   || | |    |  |
Db        345 A--LRTAARKRRRPHEPLTTSEDQQPIFATAIKAENGDDTLKAEAAEAVEIENVAVADTT 402

Qy        284 TKE-AFQNPD-----------------------------------------NQKVN 297
              | |   : ||                                        |:|
Db        403 TNEIKIEKPDTIKGEDDAERLEKEPKKAVSDDSESKEASPGQQVEPQPKDETVDVEMKMN 462

Qy        298 IDE----------LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347
                |          : ||: :| |  |: |:| :       :||  :|       ||  |
Db        463 TSEDEEPMTELPRITNAV-NGDLNGDLKASIGKPKSKPKPKAKLSSI-----IQKLIDSV 516

Qy        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTT------------ 395
              |: ::  | |:       : :  :|||:  |:|   |: : |   |
Db        517 PARLEQMSKTSAVIASTTTSSDRIGGGLSHALT---HKVSPPSSATAAGRLVEYHTQHVS 573

Qy        396 -----------------------------------TTTTTTTTSARTVENKPANNTPAQ 419
                                                :    |:||:   :  | |  | |
Db        574 PRKRILREFEKVSLEDNGCVNNGSGGASSGGAGGKRSRAKGTSTSSPAGKASPMNLAPPQ 633

Qy        420 GNVD-TPGSE---------DTMESRRSSM--------ASTSSTFFDTSSIGGPCRIRMLM 461
              |    :|||           |  :| :|          |: |    :|| |  |
Db        634 GKPSPSPGSSSSSTSPATLSTQPTRLNSSYSIHSLLGGSSGSGSSSSSSSGKKC------ 687

Qy        462 LKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYA 521
                                  |   : : | : ||           :  ||:   |:|
Db        688 ----------------------GDHPAAIISNVHHPQHS--------MYQPSS---SSYP 714

Qy        522 RLALSGGLRHDMGGLTGG-----SNSAVNTSNNPPAPGS-----HRF 558
              |  |:     |: |  ||     |::     : | : ||     | |
Db        715 RALLTSPKSPDVSGSNGGGGKSPSHTGTKKRSPPYSAGSPVDYGHSF 761


RESULT 12
DAB_DROME
ID   DAB_DROME      STANDARD;      PRT;  2411 AA.
AC   P98081;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Disabled protein.
GN   DAB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX   MEDLINE=93194063; PubMed=7680635;
RA   Gertler F.B., Hill K.K., Clark M.J., Hoffmann F.M.;
RT   "Dosage-sensitive modifiers of Drosophila abl tyrosine kinase
RT   function: prospero, a regulator of axonal outgrowth, and disabled, a
RT   novel tyrosine kinase substrate.";
RL   Genes Dev. 7:441-453(1993).
CC   -!- FUNCTION: TOGETHER WITH ABL INVOLVED IN EMBRYONIC NEURAL
CC       DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P98081-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P98081-2; Sequence=VSP_004185;
CC   -!- TISSUE SPECIFICITY: CNS AXONS AND BODY WALL MUSCLES.
CC   -!- DEVELOPMENTAL STAGE: EMBRYONIC AXONOGENESIS.
CC   -!- PTM: PROBABLY PHOSPHORYLATED BY THE ABL TYROSINE KINASE.
CC   -!- SIMILARITY: Contains 1 PID domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L08845; AAB08527.1; ALT_SEQ.
DR   PIR; A46299; A46299.
DR   FlyBase; FBgn0000414; Dab.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   InterPro; IPR006020; PTB_PID.
DR   Pfam; PF00640; PID; 1.
DR   SMART; SM00462; PTB; 1.
DR   PROSITE; PS01179; PID; 1.
KW   Alternative splicing; Phosphorylation; Repeat.
FT   DOMAIN       46    196       PID.
FT   DOMAIN     1689   1801       REPEAT-RICH REGION.
```

```
FT   REPEAT      1689   1700       ALTERNATE ARG AND ACIDIC RESIDUE.
FT   REPEAT      1740   1750       ALTERNATE ARG AND ACIDIC RESIDUE.
FT   REPEAT      1791   1801       ALTERNATE ARG AND ACIDIC RESIDUE.
FT   MOD_RES      111    111       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES      482    482       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1662   1662       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1667   1667       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1701   1701       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1704   1704       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1713   1713       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1739   1739       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1826   1826       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     1961   1961       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   VARSPLIC     462    673       Missing (in isoform Short).
FT                                 /FTId=VSP_004185.
SQ   SEQUENCE   2411 AA;  264046 MW;  579AB9C0243D5FD6 CRC64;

  Query Match             4.5%;  Score 129.5;  DB 1;  Length 2411;
  Best Local Similarity   18.8%;  Pred. No. 12;
  Matches  117;  Conservative   79;  Mismatches  228;  Indels  199;  Gaps   23;

Qy         70 PVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQRNGVET 129
              ||: :   | ::| :  |  :   |  |:     |   | ::| :    : :  |:  :
Db        114 PVHKISFIAQDMTDSRAFGYI--FGSPDS-----GHRFFGIKTDKAASQVVLAMRDLFQV 166

Qy        130 SVVLSDQE-------------------YARLQSIDPEGKDKFVFTGGRG----------- 159
                |  :|                    | | |:   |       || |
Db        167 VFELKKKEIEMARQQIQGKSLHDHSSQLASLSSLKSSG------LGGMGLGHSDLASGGI 220

Qy        160 GAGHAMVTVASDI--TEARQRI---LELLEPKGTGESKGAGES--------KGVGELRES 206
               :|||:  : | :  |    |:   |:: :  |:   : :||       : :  |:
Db        221 SSGHALTLLGSSLSTTNGTSRLGVSLDVAKASGSAAKEVSPESVADLVDLEQELTSLQRG 280

Qy        207 NSGAENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDS---- 262
              |  |  |  : :||::    |   ||         | :   :      | | |||
Db        281 ISQMERITPNEPTTSSTGGAGHPS--LAKSASEDDPFGDSFIYVPSYSILPPPPDSGRNR 338

Qy        263 --PTTTDPDAAASATE-------TATRDQLTKEAFQNPDNQKVN----IDELGNAI-PSG 308
              |    |||  |          |::         |  ||   |    :|:  :    |
Db        339 HKPPNKTPDAVTSLDAMLSPPPGTSSSHGSASAGLQAADNDDDNWLQELDQQNDVFDTSK 398

Qy        309 VLKDDVVANIEEQAKAAGEEA----KQQ-------------------AIENNAQ----- 339
              |:    : ::  |  |  |:     ||                    |: | |
Db        399 VVSSSGLGSVLAMAPLASSESTATPTQQLTEVAAGSGPLADLDIGLSTALGNEEQTSTIL 458

Qy        340 -------------------AQKKYDEQQAKRQEE---LKVSSGAGYGLSGALILGGGIGV 377
                                |||  |: |   :    ::|||:  ||   | ||   |:
Db        459 SLDAFTDLDDPLGTGRTRPYVDKKYFFQELKNPPKKLLKELSSGSQAGLGLGLSLGQLDGL 518

Qy        378 AVTAALHRKNQPVEQTTTTTTTTTTTSARTV-ENKPANNTPAQGNVDTPGSEDTMESRRS 436
                    | : || :|||| |:   |  |: :|   ||                |:
Db        519 F----------PEDSTTISTTTTTATNITAVLTNRYSNTIIAQ-------------RKK 554

Qy        437 SMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQH 496
              |: :           : :|   |  :    |    ::|      |:| | :
Db        555 SLTT--------------EMHILYYDKRVVHHFWRNFFSVQLEIALSKQLSKVCTCTAG 599

Qy        497 PPRDTTDNGARLLGNPSAGIQSTYARLAL--------------------SGGLRHDMGG 535
               |:  : |      : :| :    : :||                    |  |:  :|
Db        600 NPQQNSANTLTSTASTAASLGQLLSTVALNPDPLPAPISIPTSISHSITPSAELKLLLGH 659

Qy        536 LTGGSNSAVNTSNNPPAPGSHRF 558
              :|           ||| |  | :
Db        660 VT-----------NPPNPTGHYY 671


RESULT 13
FLID_VIBCH
ID   FLID_VIBCH     STANDARD;      PRT;   666 AA.
AC   Q9KQ63;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)

DE   Flagellar hook-associated protein 2 (HAP2) (Filament cap protein)
DE   (Flagellar cap protein).
GN   FLID OR VC2140.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF
CC       THE FLAGELLAR FILAMENT BY FACILITATING POLYMERIZATION OF THE
CC       FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CAPPING
CC       STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH
CC       THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT
CC       POLYMERIZATION AT THE DISTAL END (BY SIMILARITY).
CC   -!- SUBUNIT: Homopentamer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Flagellar.
CC   -!- SIMILARITY: BELONGS TO THE FLID FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004287; AAF95285.1; -.
DR   PIR; G82111; G82111.
DR   TIGR; VC2140; -.
DR   InterPro; IPR003481; FliD.
DR   Pfam; PF02465; FliD; 1.
KW   Flagella; Coiled coil; Complete proteome.
FT   DOMAIN      354    419       COILED COIL (POTENTIAL).
SQ   SEQUENCE   666 AA;  72330 MW;  6794EFEC34A2A0D2 CRC64;

  Query Match             4.5%;  Score 129;  DB 1;  Length 666;
  Best Local Similarity   19.6%;  Pred. No. 2.4;
  Matches  107;  Conservative   77;  Mismatches  221;  Indels  140;  Gaps   24;

Qy          1 MPIGNLGHNP--NVN---NSIPPAPPLPSQTDGAGGRGQLINSTGPLGS-RALFTPVRNS 54
              | :| :| |   ::|   : |  |  :| |       |  :  |    |  |     ::|
Db          1 MSLGPMGMNTGFDINGMVSKIVSAERVPKQQRIDNERTNIDTSISAYGRLRESLDTMKNL 60

Qy         55 MADSGDNRASDVPGLPVNPMRLAASEITLN--------DGFEVLHDH------------ 93
              |       :| |  :  :  :: ::  |           |  ::   |
Db         61 MTQFRQEKAFAVRKVDTSNEQVVSATATTEAIAGNYSVDVLQLAQSHKIASEVLDKDAKF 120

Qy         94 GPLDTLNRQIGSSVFRVETQEDGKHIAV-----GQRN--GVETSVVLSDQEYARL-QSID 145
              ||    |:  :|   | :: | : | : :      |:::  ||  |:: :| |  ||  : :
Db        121 GP-GKLHISLGDKSFTLDVQGNSKLVDIVRGINGEKSNPGVRASII-NDVEGPRLIVASN 178

Qy        146 PEGKDKFVFTGGRGGAGHAMVTV--------ASDITEARQRILELLEPKGTGESK----- 192
                |||  |     :   |: :  :           |: :|| :  :|: |    : |
Db        179 VSGKDHSVKMSAQAEPGNPLKQLEYKTLEQRVRDLEKARAQAQQLIAPLTPEQQKVAAKV 238

Qy        193 ---------------------------GAGESKGVGELRESNSGAENTTETQTSTSTSS 224
                                        |||:   ||| ||   |        | :
Db        239 AEKIGDAARLVDQEVAQEIRSAAQSAQGAAGEALNAGELTES---AVKAAANAASEAKKY 295

Qy        225 LRSDPKL--WLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAA------ASATE 276
```

```
                :| : ::  |              |:| :      || :        ||       :|:
Db        296 IRPEDRIPGWTE----------TASGTLLDSYWEPEEELDAQGQKKAADVPGWSNTASG 344

Qy        277 TATRDQLTKEAFQNPDNQKVNID--ELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAI 334
              |     :| :  |    ||:  :  ::  || ||             |   |||| ||
Db        345 TLLDSYVTPQEAQQKLEQKLAQEKAQIEAAIRSG-------------KMTPEEAKAQAR 390

Qy        335 ENNAQAQKKYDEQQAKRQEELKVSSGA--GYG--------LSGALILGGGIGVAVTAALH 384
                :  :: | ||  | |  |  :  |  |||           ::| |       | |
Db        391 AKLSPEERAYIEQVEKAQAALNAAQSAFDGYGGMTEVQSAQDSMVVLDG------VATLS 444

Qy        385 RKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSST 444
                | :|          |   :|  |:|    ||:  ::     : : :      :  :
Db        445 SNNNIIE--NAIEGVNLTLKGKTDRNQP----PAEIGIEY--DRERVRNDIEQFVAAYNQ 496

Qy        445 PFDTS 449
              || ||
Db        497 PFQTS 501
```

```
RESULT 14
SP4_HUMAN
ID   SP4_HUMAN      STANDARD;      PRT;   784 AA.
AC   Q02446; O60402;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transcription factor Sp4 (SPR-1).
GN   SP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=93087156; PubMed=1454515;
RA   Hagen G., Mueller S., Beato M., Suske G.;
RT   "Cloning by recognition site screening of two novel GT box binding
RT   proteins: a family of Sp1 related genes.";
RL   Nucleic Acids Res. 20:5519-5525(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ozersky P., Holmes A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Binds to GT and GC boxes promoters elements. Probable
CC       transcriptional activator.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: ABUNDANT IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68561; CAA48563.1; -.
DR   EMBL; AC004595; AAD12226.1; -.
DR   PIR; S26638; S26638.
DR   HSSP; P08047; 1SP1.
DR   TRANSFAC; T02339; -.
DR   Genew; HGNC:11209; SP4.
DR   MIM; 600540; -.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 3.
DR   ProDom; PD000003; Znf_C2H2; 2.
DR   SMART; SM00355; ZnF_C2H2; 3.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Nuclear protein; Repeat.
FT   ZN_FING     647    671       C2H2-TYPE 1.
FT   ZN_FING     677    701       C2H2-TYPE 2.
FT   ZN_FING     707    729       C2H2-TYPE 3.
FT   DOMAIN        7     11       POLY-GLU.
FT   DOMAIN       12     19       POLY-ALA.
FT   DOMAIN      122    130       POLY-SER.
FT   DOMAIN      185    188       POLY-SER.
FT   CONFLICT    197    197       K -> Q (IN REF. 2).
FT   CONFLICT    379    380       HA -> QP (IN REF. 2).
FT   CONFLICT    386    386       Q -> A (IN REF. 2).
SQ   SEQUENCE   784 AA;  82025 MW;  3C4EAE28CB2B81FB CRC64;
```

```
  Query Match             4.5%;  Score 129;  DB 1;  Length 784;
  Best Local Similarity   20.0%;  Pred. No. 3;
  Matches  112;  Conservative   71;  Mismatches  234;  Indels  144;  Gaps   21;

Qy          3 IGNLGHNPNVNNS---IPPAPPL------PSQ-----TDGAGGRGQLINSTGPLG-SRAL 47
              ||  | |          | |:  |       | |      |  ||    ||: || |      :
Db         58 IGTPGENQATGQQQIIIDPSQGLVQLQNQPQQLELVTTQLAGNAWQLVASTPPASKENNV 117

Qy         48 FTPVRNSMADSGDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV 107
                | :| : |  |  |        :| :  :   :    |:|:    |  ::   |:   :
Db        118 SQPASSSSSSSSSNNGS------ASPTKTKSGNSSTPGQFQVIQVQNPSGSVQYQVIPQL 171

Qy        108 FRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFT-GGRGGAGHAMV 166
              || |:        | |   :| :   |   :| |    | :: : |   |  :|: :
Db        172 QTVEGQQI-------QINPTSSSSLQDLQGQIKLIS---AGNNQAILTAANRTASGNILA 221

Qy        167 ------TVASDITEARQRILELLEPKGT---------------------------GESK 192
                   ||   |       |:|    ||                          | :
Db        222 QNLANQTVPVQIRPGVSIPLQLQTLPGTQAQVVTTLPINIGGVTLALPVINNVAAGGGTG 281

Qy        193 GAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT-------GLIGL 245
                |:     :   ||     :| | |:|| |::   |       | |:        |:
Db        282 QVGQPAATADSGTSNGNQLVSTPTNTTTSASTMPESPSSSTTCTTTASTSLTSSDTLVSS 341

Qy        246 AATGIVQALALTPEPDSPTTTDPDAAASATETATR--DQLTKEAFQNPDNQKVNIDELG- 302
              | ||     | |     |  | :     :|||:  :   ||      ||  :|  :: ::
Db        342 ADTG---QYASTSASSSERTIEESQTPAATESEAQSSSQLHANGMQNQQDQSNSLQQVQI 398

Qy        303 -------------------NAIP-------SG----VLKDDVVANIEEQA---------- 322
                                |||        ||      ::   : |:: ||
Db        399 VGQPILQQIQIQQPQQQIIQAIPPQSFQLQSGQTIQTIQQQPLQNVQLQAVNPTQVLIRA 458

Qy        323 ---KAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAV 379
                  :|: : |     | |:      | |   ::| ::  :    ||   |    |||
Db        459 PTLTPSGQISWQTVQVQNIQSLSNLQVQNAGLSQQLTITPVSS---SGGTTLAQIAPVAV 515

Qy        380 TAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGS-------EDTME 432
                |      |:  |   :       :| |  |     ||   |  |       :| ::
Db        516 AGA------PITLNTAQLASVPNLQTVSVANLGAAGVQVQGVPVTITSVAGQQQGQDGVK 569

Qy        433 SRRSSMASTSSTFFDTSSIGG 453
                :::::|        | ::||
Db        570 VQQATIAPV------TVAVGG 584
```

```
RESULT 15
PMPC_CHLMU
ID   PMPC_CHLMU     STANDARD;      PRT;  1460 AA.
AC   Q9PJY1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable outer membrane protein pmpC precursor (Polymorphic membrane
DE   protein C).
GN   PMPC OR TC0695.
```

```
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn / Nigg;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE002338; AAF39511.1; -.
DR   PIR; D81675; D81675.
DR   TIGR; TC0695; -.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR003368; Chlamydia_PMP.
DR   Pfam; PF02415; DUF145; 2.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
DR   TIGRFAMs; TIGR01376; POMP_repeat; 8.
KW   Outer membrane; Signal; Multigene family; Complete proteome.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1460       PROBABLE OUTER MEMBRANE PROTEIN PMPC.
SQ   SEQUENCE   1460 AA;  154781 MW;  C1F033E7907AC3AC CRC64;

  Query Match              4.5%;  Score 129;  DB 1;  Length 1460;
  Best Local Similarity   20.5%;  Pred. No. 6.6;
  Matches   99;  Conservative   58;  Mismatches  191;  Indels  136;  Gaps   21;

Qy         13 NNSIPPA---PPLPSQTDGAGGR--GQLINSTGPLGSRALFTPVRNSMADSGDNR--ASD 65
              |   |||   | ||:   |: |       ||:|   |     |  ::  ::|     :|
Db        276 NTVTPPAANKPELPNDPSGSNGNDGSDDSNSSGNTDSNE-SNPNNSASNNTGSENELSSS 334

Qy         66 VPG--LPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQ 123
              |   || ||     | :: |          |:||       : :  |:   |   : |:
Db        335 TPSAQLP-NPATPFLSSVSTNS--------QPIDT----EPENAWHAESGSGGAIYSKGK 381

Qy        124 RNGVETSVVLSDQEYARLQSIDPEGKDKFVF-----------TGGRGGAGHAMVTVASD 171
              :   :  |: |   |       |:::                || :||| ||     ||
Db        382 LSIASSKEVVFDHNSATKNGGAIFGEEEIALEKIASLKFDSNTTGEKGGAIHAKTVTLSD 441

Qy        172 ITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKL 231
              |        |:    | ::      |   :|   |   |   :|     : || : |
Db        442 IKNT------LIFVNNTAKTPEENSLKS-SQLNNQNPSEEEHQDTSEGEESQSLETSP-- 492

Qy        232 WLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAAS-----ATETATRDQLTKE 286
                                           |: |:|:|      :  |:  |   |
Db        493 --------------------------------ITNQDSASSHVAIFRSIAASSSQSNSE 519

Qy        287 AFQNPDNQKVNIDELGNAI-PSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYD 345
                | |         : |::  ||     |  : ::       ||  |   |||
Db        520 NIPNADGSTSAGGDAGSSSQPSTPGSDSSINHVIGGGAIYGEAVK---IENL-------- 568

Qy        346 EQQAKRQEELKVSSGAGYG-----------LSGAL--ILGGGIGVAVTAALHRKNQPVEQ 392
                          :|||          :||:   :||| |    :  :   |
Db        569 --------------SGYGTFSNNNAVDHQISGSTSDVLGGAIYAKTSLTIDSGN----- 608

Qy        393 TTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTME---SRRSSMASTSSTFFDTS 449
                  :: | | |  |  :|      | | : :|    |     |: |:: :|:|:  ||
Db        609 ----SSGTITFSENTTSSKSTTGQVAGGAIFSPSVTITTPVTFSKNSAINATTSSKKDT- 663

Qy        450 SIGG 453
                ||
Db        664 -FGG 666


Search completed: December 16, 2003, 08:53:53
Job time : 24.694 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 26.7392 Seconds
(without alignments)
2010.470 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPNVNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2624.5	92.1	558	2	A98199	translocated intim
2	2624.5	92.1	558	2	E86045	probable transloca
3	170	6.0	1229	2	T25697	hypothetical prote
4	165.5	5.8	2232	2	T34434	hypothetical prote
5	162.5	5.7	1291	2	T13389	hypothetical prote
6	152.5	5.3	971	2	B90835	probable tail fibe
7	152.5	5.3	973	2	C85693	probable membrane
8	148.5	5.2	770	2	T51024	related to C2H2 zi
9	148.5	5.2	1246	2	G89287	protein H39E23.1 [
10	148	5.2	1192	2	T18611	probable serine/th
11	146.5	5.1	1122	2	G64887	probable tail fibe
12	144.5	5.1	3507	2	T34513	hypothetical prote
13	143.5	5.0	1275	2	T33369	hypothetical prote
14	143.5	5.0	1630	2	A53577	ascites sialoglyco
15	142.5	5.0	3013	2	AB0480	probable invasin Y
16	142	5.0	563	2	A36054	mucin homolog - bo
17	141.5	5.0	1962	2	A32634	lactocepin (EC 3.4
18	140.5	4.9	1026	1	A40315	maternal effect pr
19	140.5	4.9	1829	2	T24583	hypothetical prote
20	139	4.9	2468	2	A83412	hypothetical prote
21	137	4.8	837	2	T02761	outer arm dynein i
22	137	4.8	4558	2	C82199	RTX toxin RtxA VC1
23	136	4.8	949	2	D90803	AidA-I adhesin-lik
24	136	4.8	1005	2	H85611	probable adhesin Z
25	134	4.7	2514	2	F81045	hemagglutinin/hemo
26	133.5	4.7	796	2	T21460	hypothetical prote
27	133	4.7	439	2	E71497	hypothetical prote
28	133	4.7	1589	2	T42233	submaxillary mucin
29	132.5	4.6	461	2	JN0097	secreted 45K prote
30	132.5	4.6	755	2	H86561	CT456 hypothetical
31	132.5	4.6	755	2	B72061	hypothetical prote
32	132.5	4.6	1258	2	JQ0188	ice nucleation pro
33	132.5	4.6	1459	2	T32271	hypothetical prote
34	132	4.6	1788	2	AH1447	probable tape-meas
35	132	4.6	2271	2	F90073	hypothetical prote
36	131.5	4.6	1125	2	E90598	membrane nuclease,
37	131	4.6	690	2	T31597	hypothetical prote
38	131	4.6	1083	2	JC2300	cell surface glyco
39	131	4.6	2660	2	E85822	probable invasin Z
40	130.5	4.6	829	2	D71485	probable adenylate
41	130.5	4.6	1034	2	JC2143	ice nucleation act
42	130	4.6	1063	2	D86731	hypothetical prote
43	130	4.6	1077	2	A44067	serine-rich protei
44	130	4.6	2055	2	T31110	extracellular matr
45	129.5	4.5	2059	2	D82671	surface protein XF

ALIGNMENTS

RESULT 1
A98199
translocated intimin receptor Tir [imported] - Escherichia coli (strain O157:H7, substra
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: A98199
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A98199
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-558 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37984.1; PID:g13364036; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4561

Query Match 92.1%; Score 2624.5; DB 2; Length 558;
Best Local Similarity 94.0%; Pred. No. 2.2e-148;
Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3;

```
Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60

Qy         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120

Qy        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300

Qy        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360

Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
```

```
Qy        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIG---GPCRIRMLMLKHRCMIRRCRLLILI 477
              |||||||||||||||||||||||||||||||||    |       | |   :       :
Db        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIGTVQNPYADVKTSL-HDSQVPTSNSNTSV 479

Qy        478 RLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 537
              :     |   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 QNM---GNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 536

Qy        538 GGSNSAVNTSNNPPAPGSHRFV 559
              ||||||||||||||||||||||
Db        537 GGSNSAVNTSNNPPAPGSHRFV 558


RESULT 2
E86045
probable translocated intimin receptor protein tir [imported] - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E86045
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E86045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-558 <STO>
A;Cross-references: GB:AE005174; NID:g12518449; PIDN:AAG58825.1; GSPDB:GN00145; UWGP:Z51
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: tir

  Query Match             92.1%;  Score 2624.5;  DB 2;  Length 558;
  Best Local Similarity   94.0%;  Pred. No. 2.2e-148;
  Matches  528;  Conservative    3;  Mismatches   24;  Indels    7;  Gaps    3;

Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60

Qy         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120

Qy        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300

Qy        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360

Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420

Qy        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIG---GPCRIRMLMLKHRCMIRRCRLLILI 477
              |||||||||||||||||||||||||||||||||    |       | |   :       :
Db        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIGTVQNPYADVKTSL-HDSQVPTSNSNTSV 479

Qy        478 RLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 537
              :     |   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 QNM---GNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLT 536

Qy        538 GGSNSAVNTSNNPPAPGSHRFV 559
              ||||||||||||||||||||||
Db        537 GGSNSAVNTSNNPPAPGSHRFV 558


RESULT 3
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T25697
R;Fulton, B.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid F16F9.
A;Reference number: Z20071
A;Accession: T25697
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1229 <FUL>
A;Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2
A;Experimental source: strain Bristol N2; clone F16F9
C;Genetics:
A;Gene: CESP:F16F9.2
A;Map position: X
A;Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

  Query Match              6.0%;  Score 170;  DB 2;  Length 1229;
  Best Local Similarity   22.8%;  Pred. No. 0.032;
  Matches   68;  Conservative   44;  Mismatches  122;  Indels   64;  Gaps    9;

Qy        167 TVASDITEARQRILELLEPKGTGESKGAGESKGVGELR----ESNSGAENTTETQTSTST 222
              :|::: |   ::     ||  | |     ||   |:       : |    ||| : :|
Db        597 SVSTEKTSTTKKASTTEEPTTTDEPTTTTESSTTGKATTPELSTTSEETTTTELKITTEG 656

Qy        223 SSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQ 282
              |:   :|           | :   |:|||:     | |  : ||:       |  | |
Db        657 STTTEEP--------TTTAIFAEASTGII----TTDEETTSTTSTTPEITSTKEIVTESA 704

Qy        283 LTKEAF---------QNPDNQKVNIDELGNAIPSGVLKDDV-VANIEEQAKAAGEEAKQQ 332
              :|: :          | |:  |  :::  : :  |||:   :   :|     | ::
Db        705 ITQTSVSVVESSTPRQLPERWKAIVNKFKHNLE--VLKEKKRLLKEKESTSTTGSDSS-- 760

Qy        333 AIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQ 392
              |     :  ||    :|::                       |  |     |:   |:
Db        761 --ETTTVVAENIDEVTTTEKEKV--------------------VQTTPITTEKSTTQEE 797

Qy        393 TTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
              ||||||||  |:::|   ||  :  |            | |:  |  ::| ||  ||||
Db        798 TTTTTTTTEKTTSKTTTEKPTTSESA----------TTETTTSEPSTTESTTVDTSS 844


RESULT 4
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

  Query Match              5.8%;  Score 165.5;  DB 2;  Length 2232;
```

```
  Best Local Similarity   19.3%;  Pred. No. 0.13;
  Matches  113;  Conservative   81;  Mismatches  230;  Indels  161;  Gaps   20;

Qy          9 NPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSM---------- 55
              :|:  : :| : |:|| |  :   |  |:    : : :|  |: :|:
Db       1343 SPSSISPVPTSSPIPSTTFASSTSGSTISDVSSVSTTSL-APLSSSLPSTVPSSTQSFSS 1401

Qy         56 ADSGDNRASDVP------GLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFR 109
                | ::|| |        | ||     |   |:        |    |:::    |
Db       1402 TSEGSSKASSSPVPSQTSSTPTNPTGSTESSTLLSSTISGSTQH---TTMSKASSGSTSP 1458

Qy        110 VETQEDGKHIAVG--QRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHA--- 164
                : |  : :|    :|| ||   | |           |       |     :  |
Db       1459 STNSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTASPAASSTAPSS 1518

Qy        165 --------MVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGAENTTET 216
                     || | |:|:          |:  : |:  : ||      :: : ::|:
Db       1519 TGTMSSTSSGTVGSTISES-STTASASSQTGSTVTMGSSSTSGV------STSSASSTQP 1571

Qy        217 QTSTSTSSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATE 276
              | |||  |          |  |||:   || :|  |         | |  |    :: :
Db       1572 QMSTSQGS--------SAGSTVASSTAGLVSTSTV--------PSSTGTMGSTSSGTVGS 1615

Qy        277 TATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIEN 336
              | :    | |     :|  :   :|::  |||
Db       1616 TISESSTTASA----SSQTGSTVTMGSSSTSGVSTSSA---------------------- 1649

Qy        337 NAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTT 396
                       :  | ::  | |:    :|: :     |:  |: :           ::
Db       1650 -----------SSTQPQMSTSQGSS---AGSTVASSTTGLVSTSTV----------PSS 1684

Qy        397 TTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCR 456
              | |  :||: || :  : :: |        ||  || |  :|  ||||     :| | |
Db       1685 TGTMGSTSSGTVGSTISESSTAASASSQTGSTVTMGSSSTSGVSTSS-----ASSGQP-- 1737

Qy        457 IRMLMLKHRCMIRRCRLLILIRLFRIWGIQIS----------VVYSTIQHPPRDTTDNGA 506
                                          |:|          || ||       |  :
Db       1738 ----------------------------QMSTSQGSSAGSTVVSSTASPAASSTAPSST 1768

Qy        507 RLLGNPSAG-IQSTYARLALSGGLRHDMGG-LTGGSNSAVNTSNN 549
                :. : |:| : || :: : :       |  :| ||:|   ||:|
Db       1769 GTMSSTSSGTVGSTMSQSSTAASTTSHTGSTVTLGSSS---TSSN 1810
```

RESULT 5
T13389
hypothetical protein 115C2.10 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13389
R;Salles, C.; Valenti, P.; Darlamitsou, A.; Henderson, N.; Campbell, L.; Glover, D.
submitted to the EMBL Data Library, May 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17665
A;Accession: T13389
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <CAT>
A;Cross-references: EMBL:AL031581; NID:e1320978; PID:e1426292; PIDN:CAA20894.2
C;Genetics:
A;Cross-references: FlyBase:FBgn0020381
A;Map position: X
A;Introns: 238/3; 1225/1
A;Note: EG:115C2.10

```
  Query Match             5.7%;  Score 162.5;  DB 2;  Length 1291;
  Best Local Similarity   20.6%;  Pred. No. 0.095;
  Matches   99;  Conservative   61;  Mismatches  189;  Indels  131;  Gaps   18;

Qy          5 NLGHNPNVNNSIPPAPPLPSQTDG---AGGRGQLINSTGPLGSRALFTPVRNSMADSGDN 61
              |   | | | :    |   | |:|   :|| |       | |:  | || |   :|
Db        438 NSTSNSNSNTNDSTGPSETSSTNGLVASGGAG------GATGAAMLPTP---SQQSTGGK 488

Qy         62 RAS--------DVPGLPVNPMRLA----------ASEITLNDGFEVLHDHGPLDTLNRQ 102
               |:          :| : |:|: :           | |  |  ::  | |
Db        489 EATAAVSLLEKKLPNVVVSPLTMKELRQKGMTKYDAEMIMANAAYQQQHHH--------- 539

Qy        103 IGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAG 162
                    |          |   ||          : |: | :|            || | ||
Db        540 --QHHFHHHHHHHHHHHNHGQHASTGAEATAAVQQMAAMQKPG---------VGGTGAAG 588

Qy        163 HAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTST 222
              :|  |  | :                | |      | |  : :    ||          ||:|
Db        589 NAGATTVSSVA-----------AGAGSEVNGGRSTSLRKSMRVNS---------TSSSI 627

Qy        223 SSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQ 282
              |:  :|  :                | | :::    :|    |    :    |   |
Db        628 STASADEVI----------------APVVAASISLPSKAPVVLMPRCKPAQMAIAALHQ 670

Qy        283 LTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQ---------A 333
                :   :  : ||  : :       |   |    : |:| |    | |||         :
Db        671 SQQRQLRRSERQKEKLTD-------GESSD---TSSEQQKK----EQKQQDHQLPQKMFS 716

Qy        334 IENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQT 393
              :   | :|  :|:| ::|: :  :|    ||           ||  |  |  |
Db        717 LAEEPQPEKS-EEKQQEQQKRVTRNSAGRVGL-----------VARLATAHNNN----IA 760

Qy        394 TTTTTTTTTTSARTV---ENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
              ||| :::::  | |:    |  :||:    :     |  :::||: : :  ||    |||
Db        761 TTTNSSSSSNKATTITNCNNHNSNNSSRINHNSNLSSRLSVKSRKPAPSEAASSIPSSTSS 820
```

RESULT 6
B90835
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90835
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35073.1; PID:g13361114; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1650

```
  Query Match             5.3%;  Score 152.5;  DB 2;  Length 971;
  Best Local Similarity   22.1%;  Pred. No. 0.25;
  Matches   87;  Conservative   55;  Mismatches  169;  Indels   83;  Gaps   14;

Qy         79 SEITLNDGFEVLH--------DHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQ------- 123
              | | | :||   |    .   |  |  |||  :|:      |::| :  |: :
Db         59 SVILLVEGFPPSHAGTITVYEDSQP-GTLNDFLGAM-----TEDDVRPEALRRFELMVEE 112

Qy        124 --RNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILE 181
                 ||    | |  :   |:  : |          |  |  | ||  | |:|    |
Db        113 VARN---ASAVAQNTAAAKKSASDAS-------TSAREAATHA--TDAADSARA------ 154

Qy        182 LLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATG 241
                      |  ||::    :   |::|  :|  |: | | ::  |         : |
Db        155 --------ASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSK-------SAAAT 199

Qy        242 LIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDEL 301
                | | |    |       :  :|    |: |  :|     :||| :: :    :
Db        200 SAGAAKTSETNAAVSQQSAATSASTATTKASEAASSARDASASKEAAKSSETSAAS---- 255

Qy        302 GNAIPSGVLKDDVVANIEEQAKAAGEEAK-QQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
                          ::    | |||  ||  :  | ||::: :   || |       |  :
Db        256 -----------SASSAASSATAAGNSAKAAKTSETNAKSSETAAEQSASAAAGSK--TA 301
```

```
Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQ- 419
Db        302 AALSASAASTSAGQASASATAA----GKSAESAASSASTATTKAGEATEQASAAASSASA 357

Qy        420 ---GNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
Db        358 AKTSETNAKASETSAESSKTAAASSASSAASSAS 391


RESULT 7
C85693
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strai
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85693
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-973 <STO>
A;Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:Z19
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1918

  Query Match             5.3%;  Score 152.5;  DB 2;  Length 973;
  Best Local Similarity   22.1%;  Pred. No. 0.25;
  Matches   87;  Conservative   55;  Mismatches  169;  Indels   83;  Gaps   14;

Qy         79 SEITLNDGFEVLH--------DHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQ------- 123
Db         61 SVILLVEGFPPSHAGTITVYEDSQP-GTLNDFLGAM-----TEDDVRPEALRRFELMVEE 114

Qy        124 --RNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILE 181
Db        115 VARN---ASAVAQNTAAAKKSASDAS-------TSAREAATHA--TDAADSARA------ 156

Qy        182 LLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATG 241
Db        157 --------ASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSK-------SAAAT 201

Qy        242 LIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDEL 301
Db        202 SAGAAKTSETNAAVSQQSAATSASTATTKASEAASSARDASASKEAAKSSETSAAS---- 257

Qy        302 GNAIPSGVLKDDVVANIEEQAKAAGEEAK-QQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
Db        258 ------------SASSAASSATAAGNSAKAAKTSETNAKSSETAAEQSASAAAGSK--TA 303

Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQ- 419
Db        304 AALSASAASTSAGQASASATAA----GKSAESAASSASTATTKAGEATEQASAAASSASA 359

Qy        420 ---GNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
Db        360 AKTSETNAKASETSAESSKTAAASSASSAASSAS 393


RESULT 8
T51024
related to C2H2 zinc finger transcription factor D-Sp1 [imported] - Neurospora crassa
N;Alternate names: protein B7F21.50
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51024
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51024
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-770 <SCH>
A;Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.50
A;Experimental source: BAC clone B7F21; strain OR74A
C;Genetics:
A;Gene: NCSP:B7F21.50
A;Map position: 6
A;Introns: 117/1

  Query Match             5.2%;  Score 148.5;  DB 2;  Length 770;
  Best Local Similarity   19.9%;  Pred. No. 0.32;
  Matches  109;  Conservative   62;  Mismatches  151;  Indels  227;  Gaps   25;

Qy          5 NLGHNPN-VNNSIPPAPPLPSQT---DGAGGRGQLINST---GPLGSR---------ALF 48
Db        219 NSGAIPKFLPSSVPKASAIPKDTKDSDGSGG----LNGAQLGGIIGGAIALLIIVIIAAF 274

Qy         49 TPVR--NSMADSGDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSS 106
Db        275 LIIRRLKRVESAMESKKGSTSGYHSKASKTSASQAQMEQSGRFLHVRAPSDTDNASADPL 334

Qy        107 VFRVETQEDGKHI----------------------AVGQ--RNGVETSVVLS------DQ 136
Db        335 MFMSETNTPGDHTTNASSLAGTPQPGAHGVVGDGSGIGRHGRSGSDTTYMASPHGAGNNN 394

Qy        137 EYARLQSIDP-----EGKD----------KFVFTGGRGGAGHAMVTVASDITEARQRILE 181
Db        395 NGSELASPDPNRGYFDGASPPLPSSSHTTSTTYAGGHNGNGHGM----------RESV-- 442

Qy        182 LLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGT---- 237
Db        443 ------------DSQSTGLGYHYSSTTRNQHQHHWRNQSNASELSAD-------GSEITH 483

Qy        238 -VATGLIG-----LAATGIV--------------QALALTPEPDS--------PTTTDPD 269
Db        484 GVASPLVGGSSHARGASGGTSYRYTHTHTHSHSHSGLGDVPELDSSGMFVELPATTTTTT 543

Qy        270 AAASAT---------ETATRDQLTKEAFQNPDN----QKVNIDELGNAIPSGVLKDDVVA 316
Db        544 ASASASLPPRNSFGLSSASRRRRTTGGGSNPNTSSTPQSPNTD---------------- 586

Qy        317 NIEEQAKAAGEEAKQQAIENN----------------------------AQAQKKYDEQQ 348
Db        587 STQQQQQQQQQQQQQQQQQNNLVDVRSEPETAEFITRPHHQQQQQQQQQQQQQQQQQQQ 646

Qy        349 ----------AKRQEELKVSSG------------------------AGYGL----SGAL 369
Db        647 HTSLDGYYGPAKRQVGQTAASGLGFVPERWDGSSPGAYPMGFPMISVPGYGLGVSSSGGV 706

Qy        370 ILGGGIGVA 378
Db        707 SSGGGGGPA 715


RESULT 9
G89287
protein H39E23.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: G89287
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: G89287
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1246 <STO>
A;Cross-references: GB:chr_V; PIDN:CAB09532.1; PID:g3878100; GSPDB:GN00023; CESP:H39E23.
C;Genetics:
```

```
A;Gene: H39E23.1
A;Map position: 5

  Query Match             5.2%;  Score 148.5;  DB 2;  Length 1246;
  Best Local Similarity   18.6%;  Pred. No. 0.62;
  Matches  124;  Conservative   78;  Mismatches  233;  Indels  233;  Gaps   25;

Qy         36 INSTGPLGSRALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGP 95
              || :  ||         |  :       :|   |   :| | : |  |   |  :
Db        395 INVSSSLGQHPAGVITREHVT------SSSASGSSASPSRYSRSSATAT-GASITAGSAL 447

Qy         96 LDTLNRQ----------IGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSID 145
                 | |           |||  |  :| |     | |        :||:|
Db        448 ASAANAQKHQQSSAAPSSGSSSSRRSSQNDAAATAAG------GTVVMS----------- 490

Qy        146 PEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRE 205
                        |  |  | : :  | :||  : ||:|     |    :   :  |
Db        491 --------------GTRHGGVQMRAQPT-SRQATISLLQPPSYKPSSNTTQIAQIPPLFN 535

Qy        206 SNSGAENTTETQTSTSTSSLR--SDPKLWLALGT-------VATGLIG------------ 244
              || |  ::  | ||  :  |  :|||  : | :         ||| :
Db        536 RNSTA-TSSAAQPSTGITGTRKIADPKGRIPLNSTAVQGHRTATGAVAANNGGIPSHRDH 594

Qy        245 ----------LAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQ----- 289
                       ::| : : :  ||       |   :::||| ||     | |   |
Db        595 AQQQQYMNQLTSSTMMSKLINKTPAAGGTAATSSSSSSSATSTA---PLQKSGSQISHAP 651

Qy        290 ------------NPDNQKVNIDELGNAIPS-----GVLKDDVVANI---EEQAKAAGEEA 329
                         | :||  |:  :|   |       | :|   ::    ::| ||: |
Db        652 TEPVIREDDDENNSENQNGNVPLIGGVGPQTSPAVQVPTEDATSSSDKEQQQQKASSETP 711

Qy        330 KQ------QAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAAL 383
              |:      | :  |:  :   |   |   |  :    ||
Db        712 KESNPIVWQNLHLNSLLKSLLDSSAATSYETPRRPGIAG---------------------- 750

Qy        384 HRKNQPVEQTTTTTTTTTTTSARTVENKPANNTP-----------------------AQ 419
               |:::|   |       |    || ::  |  : |                      :
Db        751 -RRSEPSAATPRRRHQTMVVDARHLQTPPDTDRPYHFEDTTLDRQMRALYVSTASSRMTR 809

Qy        420 GNVDTPGSEDTMESR--------------------RSSMASTSSTFFDTSSIGGPCRIR 458
              | : || : ::  |                    |  :||  |  | |       |
Db        810 GVLPTPPTSNSTSSSFIVEPLTHVAAASPDITTTTPTKSTVTTSPYFRRTPS------FR 863

Qy        459 MLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPP------------RDTTDNGA 506
              |:::          ||:     |:|    |::: :   ||              ::   |
Db        864 MVLIV---------LLLCDNGLRLWP---SMIHQSPSMPPSQMMTAMESLKLSESGQTGG 911

Qy        507 RLL---GNPSAGIQSTYARLALSGGLRHDMGGLTGGSNSAVNTSN--------------- 548
                :   | |          :| | :     :||  :||: :| : :|
Db        912 PTVATGGPPQRATSQQMSRSATTNSA--NMGASSGGAAAAASATNQLSGAPSSTGASSQQ 969

Qy        549 -NPPAPGS 555
               :| || |
Db        970 YHPKAPSS 977


RESULT 10
T18611
probable serine/threonine-specific protein kinase (EC 2.7.1.-), long splice form - Caenc
N;Contains: probable serine/threonine kinase, short splice form
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T18611; T18610; T23144; T23143
R;McMurray, A.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z18997
A;Accession: T18611
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1192 <WIL1>
A;Cross-references: EMBL:Z81027; PIDN:CAB54179.1; GSPDB:GN00023; CESP:H39E23.1a
A;Experimental source: clone AH10
A;Accession: T18610
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-487,536-1192 <WIL2>
A;Cross-references: EMBL:Z81027; PIDN:CAB54178.1; GSPDB:GN00023; CESP:H39E23.1b
A;Experimental source: clone AH10
R;McMurray, A.
submitted to the EMBL Data Library, June 1997
A;Reference number: Z19696
A;Accession: T23144
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1192 <WIL3>
A;Cross-references: EMBL:Z96102; PIDN:CAB54263.1; GSPDB:GN00023; CESP:H39E23.1a
A;Experimental source: clone H39E23
A;Accession: T23143
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-487,536-1192 <WIL4>
A;Cross-references: EMBL:Z96102; PIDN:CAB54262.1; GSPDB:GN00023; CESP:H39E23.1b
A;Experimental source: clone H39E23
C;Genetics:
A;Gene: CESP:H39E23.1a; CESP:H39E23.1b
A;Map position: 5
A;Introns: 10/3; 33/3; 67/3; 139/2; 183/3; 451/3; 487/3; 535/3; 631/1; 825/2; 914/3; 992
C;Keywords: alternative splicing; ATP; phosphotransferase; serine/threonine-specific pro
F;1-1192/Product: probable serine/threonine-specific protein kinase, long splice form #s
F;1-487,536-1192/Product: probable serine/threonine-specific protein kinase, short splic

  Query Match             5.2%;  Score 148;  DB 2;  Length 1192;
  Best Local Similarity   20.6%;  Pred. No. 0.62;
  Matches  117;  Conservative   56;  Mismatches  248;  Indels  148;  Gaps   20;

Qy         36 INSTGPLGSRALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGP 95
              || :  ||         |  :       :|   |   :| | : |  |   |  :
Db        505 INVSSSLGQHPAGVITREHVT------SSSASGSSASPSRYSRSSATAT-GASITAGSAL 557

Qy         96 LDTLNRQ----------IGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSID 145
                 | |           |||  |  :| |     | |        :||:|
Db        558 ASAANAQKHQQSSAAPSSGSSSSRRSSQNDAAATAAG------GTVVMS----------- 600

Qy        146 PEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRE 205
                        |  |  | : :  | :||  : ||:|     |    :   :  |
Db        601 --------------GTRHGGVQMRAQPT-SRQATISLLQPPSYKPSSNTTQIAQIPPLFN 645

Qy        206 SNSGAENTTETQTSTSTSSLR--SDPKLWLALGT-------VATGLIG------------ 244
              || |  ::  | ||  :  |  :|||  : | :         ||| :
Db        646 RNSTA-TSSAAQPSTGITGTRKIADPKGRIPLNSTAVQGHRTATGAVAANNGGIPSHRDH 704

Qy        245 ----------LAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQ----- 289
                       ::| : : :  ||       |   :::||| ||     | |   |
Db        705 AQQQQYMNQLTSSTMMSKLINKTPAAGGTAATSSSSSSSATSTA---PLQKSGSQISHAP 761

Qy        290 ------------NPDNQKVNIDELGNAIPS-----GVLKDDVVANI---EEQAKAAGEEA 329
                         | :||  |:  :|   |       | :|   ::    ::| ||: |
Db        762 TEPVIREDDDENNSENQNGNVPLIGGVGPQTSPAVQVPTEDATSSSDKEQQQQKASSETP 821

Qy        330 KQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQP 389
              |:       |:       |    | ||:|     |  |  :  ||     |:      |
Db        822 KESKPSMIHQSPSMPPSQMMTAMESLKLSESGQTG--GPTVATGGPPQRATS-----QQM 874

Qy        390 VEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTS 449
              || :     :|        | |  :     |  |       :  | :|:|||
Db        875 SRSATTNSANMGASSGGAAAAASATNQLSGAPSSTGASSQQYHPKAPSSSSSST----- 929

Qy        450 SIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPPR---DTTDNGA 506
              |      | |                    ||  |: | |     : |  |
Db        930 --NPP---HQHQLTHNA-----------------SFSVTPSSYQIPTSTAVNVTSTGM 965

Qy        507 RLLGNPSAGIQSTYARLALSGGLRHDMGG 535
                 : ||  ::|  |    |    | ||
Db        966 PTSSSSSAFPRNTRNRQTFHGKTEKDKGG 994
```

```
RESULT 11
G64887
probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: G64887; T09189
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64887
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1122 <BLAT>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636;
A;Experimental source: strain K-12, substrain MG1655
R;Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.
.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Saito,
moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A;Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A;Reference number: Z16603; MUID:97251357; PMID:9097039
A;Accession: T09189
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 3-1122 <AIB>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636;

  Query Match             5.1%;  Score 146.5;  DB 2;  Length 1122;
  Best Local Similarity   23.1%;  Pred. No. 0.7;
  Matches   92;  Conservative   52;  Mismatches  163;  Indels   91;  Gaps   17;

Qy          79 SEITLNDGFEVLH--------DHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQ------- 123
               | | | :||   |       |  |  |||  :|:      |::| :  |: :
Db          61 SVILLVEGFPPSHAGTITVYEDSQP-GTLNDFLGAM-----TEDDARPEALRRFELMVEE 114

Qy         124 --RNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILE 181
                 ||    | |  :   |:  : |          |  |  | ||    |:|   |
Db         115 VARN---ASAVAQNTAAAKKSASDAS-------TSAREAATHA--ADAADSARA------ 156

Qy         182 LLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATG 241
                       | ||::    :   |::|  :|  |: | | ::  |          : |
Db         157 --------ASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSK-------SAAAT 201

Qy         242 LIGLAATGIVQALA----LTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVN 297
                 | | |    | |         : ||   :|| || : |     :||| :: :
Db         202 SAGAAKTSETNASASLQSAATSASTATTKASEAATSARDAAA----SKEAAKSSET---- 253

Qy         298 IDELGNAIPSGVLKDDVVANIEEQAKAAGEEAK-QQAIENNAQAQKKYDEQQAKRQEELK 356
                 ||  |         ::    | |||  ||  :  | ||:: :    | |       |
Db         254 -----NASSS-------ASSAASSATAAGNSAKAAKTSETNARSSETAAGQSASAAAGSK 301

Qy         357 VSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNT 416
                 : |    | |    |    : |||     :  |   :: :| || :    |   |
Db         302 --TAAASSASAASTSAGQASASATAA----GKSAESAASSASTATTKAGEATEQASAAAR 355

Qy         417 PAQ----GNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
                |         :   || :.|| ::: ||::|:   ::|
Db         356 SASAAKTSETNAKASETSAESSKTAAASSASSAASSAS 393


RESULT 12
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34513
R;Pavello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3507 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
 3504/1

  Query Match             5.1%;  Score 144.5;  DB 2;  Length 3507;
  Best Local Similarity   19.7%;  Pred. No. 4.3;
  Matches  105;  Conservative   61;  Mismatches  179;  Indels  189;  Gaps   22;

Qy          27 DGAGGRGQLINSTGPLGSRALFTPVR---NSMADSGDNRASDVPGLPVNPMRLAASEITL 83
               :|:|    || ||  ||     :|    :: : |:   :  ||:   | :      |
Db         728 EGSGSWSTTINGTGITGSPRSEGTIRVRITTLGEDGEPETATKPGISA-PDKTGEGSKTE 786

Qy          84 NDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQS 143
               :|| |                    ::  ::|||     | :|               |
Db         787 SDGEE-------------------KLTVEKDGKE---AQSSG--------------SS 808

Qy         144 IDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKGTGE----SKGAGESKG 199
                 ||     :|    :  :                          |||     | ||  | |
Db         809 ATSSGKKSEATSGSSSSSAKS---------------------GTGSEASGSSGASSSSG 846

Qy         200 VGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPE 259
                |  || |          | || |          : || ::| :|  |||
Db         847 SGVSGESGS----------SVSTES--------GSFGTSSSGSVGSEATG---------- 878

Qy         260 PDSPTTTDPDAAASATETATRDQL--TKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVAN 317
                 |   |   :   ::::| ::|  ||   ::|                        ::
Db         879 ---STGVDGSESGKPSKSSTEEKLPFTKNGEKSP----------------------ISG 912

Qy         318 IEEQAKAAGEE-AKQQAIENN-----AQAQKKYDEQQAKRQEE----LKVSSGAGYGLSG 367
                 :   | : ||   :: || :          :: |  :|    |    : |:|| |   ||
Db         913 SDTTGKESSEETTSRKPIEGSDSLTEGSGSGEWFETGSKGHFESGSKVSVTSGKGPTQSG 972

Qy         368 ALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPAN-----NTPAQG-- 420
                |  | | |  |          | ||    :::|::       |||:     | |  |
Db         973 A--EGSGSGPKVPKG---PGAP-EITTDGEESSSTSTGDKSGGKPADKSDNKNVPKTGDK 1026

Qy         421 --NVDTPGSEDTMESRRSSMA-----------STSSTFFDTSSIGGPCRIRMLMLKHRC 466
                 :: | | : | |:                |     || : ||  :    |
Db        1027 NPDITTDGEDSTSETSGGEQGPKGKSKGQPPGDKGSEVKKPTSEVDGPGNLSGTKGK--- 1083

Qy         467 MIRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTY 520
                                          | :   | |  : |: :|   | |  ||:
Db        1084 ----------------------SNVPLKPTDLPEEGSGILTTSSGGKNSTF 1112


RESULT 13
T33369
hypothetical protein H02F09.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C;Accession: T33369
R;Geisel, C.; Harmon, G.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid H02F09.
A;Reference number: Z21330
A;Accession: T33369
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1275 <GEI>
A;Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
A;Experimental source: strain Bristol N2; clone H02F09
C;Genetics:
A;Gene: CESP:H02F09.3
A;Map position: X
A;Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
```

```
  Query Match             5.0%;  Score 143.5;  DB 2;  Length 1275;
  Best Local Similarity   18.6%;  Pred. No. 1.3;
  Matches   95;  Conservative   66;  Mismatches  214;  Indels  137;  Gaps   15;

Qy         26 TDGAGGRGQLINSTGPLGS----------RALFTPVRNSMADSGDNRASDVPGLPVNPM 74
Db        764 TDGSTVSG----STGSTGSTNNPGSTDSSTTGISTVSGSSLSTISGSTGSTVSG------ 813

Qy         75 RLAASEITLNDGFEVLHDHGP---------LDTLNRQIGSSVFRVETQEDGKHIA----- 120
Db        814 ---SSDMTVSTG----STSSPGSTESTVSGASTMSPSTGSS---VETSTSGSSVSTVSQS 863

Qy        121 -----VGQRNGVETSVVLSDQE-------------------------------YARLQSI 144
Db        864 TSSSTTGQSTVSESSVSTVSSESTISQSTGSTTTGESTVFGSTGSTATGSSTMSASTGST 923

Qy        145 DPEGKDKFVFTGGR-------GGAGHAMVTVASDITEARQRILELLEPKGTGESKGA--- 194
Db        924 DTPGSTESTITGSTVTGESTVSGSTGSTITEGSTISESTMTTV--------GVSTGSTIT 975

Qy        195 GESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAATGIVQAL 254
Db        976 GESTVSGSTRSTVTGESTVSGSTESTVSGSTESTP----TVPSTVSGSTGSTVTGESTVS 1031

Qy        255 ALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKV----NIDELGNAIPSGVL 310
Db       1032 GSTASTSSGSTGSSTEAGSTVSGSSASTVTSSTGSSTSGESTVSGSTVSTVSGSTGSTIT 1091

Qy        311 KDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALI 370
Db       1092 GESTVSGSTESTVTAESTVSGSSVST--------------------VSGNTGSTITGEST 1131

Qy        371 LGGGIG----------VAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
Db       1132 VSGSTGSTGESTILESSVSTVSVSTGSTITDGSTASRSSVSTVSASTESTVSGGSSASIG 1191

Qy        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIG 452
Db       1192 STNTPDSTESTISGSTISGSTGSTESSTMSAG 1223


RESULT 14
A53577
ascites sialoglycoprotein 1 - rat (fragments)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 07-Feb-1997
C;Accession: A53577
R;Wu, K.; Fregien, N.; Carraway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A;Title: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifun
A;Reference number: A53577; MUID:94216302; PMID:8163496
A;Accession: A53577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1630 <WUA>
A;Cross-references: GB:U06746
C;Keywords: glycoprotein

  Query Match             5.0%;  Score 143.5;  DB 2;  Length 1630;
  Best Local Similarity   19.5%;  Pred. No. 1.8;
  Matches  145;  Conservative   66;  Mismatches  267;  Indels  267;  Gaps   29;

Qy          8 HNPNVNNSIPPAPPLPSQ----------TDGAGGRGQLINSTG--PLGSRALFTPVRNSM 55
Db        595 HAPRESSS----PPSTSDILTTMASTEGTSGDTGHTMAVTTQGSTPATTEISVTPSTQKM 650

Qy         56 ADSGDNRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQED 115
Db        651 S------------PVSTFSTSTQEITTLS--QSQHTGGMKTTRNPQTTGTTEVTTTLSA 695

Qy        116 GKHIAVGQRNGVETSV---VLSDQEYARLQSIDPEGKDKFVFT----GGRGGAGHA---- 164
```

```
Db        696 SSSDQVQAETSSQTTLSPDTTTTSHAPRESSSPPSTSDMLTTTASTEGTSGDTGHTTAVT 755

Qy        165 ------------------MVTVASDITEARQRILELLEPKGTG---------------- 189
Db        756 TQGSIPATTQLSTTFASQKMSTVSTPTTSSIQELSTLPQSQHTGSMEISSRPQTTSVTST 815

Qy        190 -ESKGAGE------------------SKGVGELR-------------ESNSG--AENTT 214
Db        816 LSSSPSGSTPVQTRSVTSSSDERTNPTSSGVSNTSPATTEVLTPTSSPESTPGNTAPRTT 875

Qy        215 ETQTSTSTSSLRSDPKLWLALG--------TVATGLIGLAATGIVQALALTPEPDSPTTT 266
Db        876 ETSTTTTTKVLMTSLQQKLPTGSTLGTSTPTEVTTTLSASSSDQVQV-----ETTSQTTL 930

Qy        267 DPDAAA----------------------SATE---------TATRDQ----LTKEAFQNPD 292
Db        931 SPDATTTSHAPRESSSPPSTSVILTTMASTEGTSGDTGHTTAVTDQGSTPATTEISVTPS 990

Qy        293 NQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQ 352
Db        991 TQKMS--TVSTLVTS---TQELTSSQSQRTGSMGTSSK----------------PQATTP 1029

Qy        353 EELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPV----------------------E 391
Db       1030 TEVTTSTLSSFS-------RGSLFSARNCCLQTKKPPLPAVVCLPDPSSVPSLMHSSKPQ 1082

Qy        392 QTTTTTTTTTTTSA----RTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFD 447
Db       1083 ATTPTEVTTSTLSSFSRGSTQTQTVSWETSSSGKITAPST----SSRRTPSVATSDIFTT 1138

Qy        448 TSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQH---PPRDTTDN 504
Db       1139 TDSTSGNAGHTLLTGSHSVITSRVASTTLGRL-----------STVAHRQSTQRSSTHS 1186

Qy        505 GARLLGNPSAGIQS---------TYARLAL---------------SGGLRHDMGG----- 535
Db       1187 QSYLTESMGASSTSETSLLTEATTETSLCLFTWTHCDRDLLSWTSSSGLTTKTDNDRSTA 1246

Qy        536 -------LTGGSNSAVNTSNNPPAP 553
Db       1247 LSATSLTLPAPSTSTASRSTVPPAP 1271


RESULT 15
AB0480
probable invasin YPO3944 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AB0480
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0480
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3013 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93406.1; PID:g15981852; GSPDB:GN00175
C;Genetics:
A;Gene: YPO3944

  Query Match             5.0%;  Score 142.5;  DB 2;  Length 3013;
  Best Local Similarity   22.3%;  Pred. No. 4.6;
  Matches   92;  Conservative   52;  Mismatches  146;  Indels  123;  Gaps   17;

Qy        103 IGSSVFRVETQEDGK------HIAVGQRNGVETSVVLSDQEYARLQSIDPEGK------- 149
Db       1570 VGSST--VTTNLDGRVRASATHTVAGSSNTV-IAISGAHQGYARVTFVADVSTAQLKLTS 1626

Qy        150 --DKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKGTGESKG-AGESKGVGELRES 206
```

```
Db       1627 FLDNQLANGKAGNIAQALVTDAHDNLLANQSVSFALDNGAVIESQGDASSASGIVLMRFN 1686

Qy        207 NSGAENTTETQTSTSTSSLRSDPKLWLA-----------------------LGTVATGL 242

Db       1687 NTLAGMTTVTATLDSTGQTETLETHFVAGKAASIEMTMTKDNAVANNIDTNEVQVLVTDV 1746

Qy        243 IGLAATGIV------QALALTPEPDSPTTTDPDAAASATETATR----------DQLTKE 286

Db       1747 DGNAINGAVVNLTSNSGMNITP---NSVTTGSDGTATATLTHTLAGSLPINARIDQVSK- 1802

Qy        287 AFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQ--------------AKAAGEEAKQQ 332

Db       1803 --------TINATFIADASTAQIIAGDMFIIVNDQVANGQAVNAVQARVTDSYGNPIKDQ 1854

Qy        333 AIE----NNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALI-----LGGGIGVAVTAAL 383

Db       1855 TVEFVLSNNGTI-----------QYELDVTSVEG----GVMVTFTNTLAGITNVTATVVS 1899

Qy        384 HRKNQPVEQTTTTTTTTTTTSAR----TVENKPANN----------TPAQGNV 422

Db       1900 SGSSRNIDTTFIADVTTAHIAASDLMVIVDDAVADNLDKNEVHARVTDAKGNV 1952
```

Search completed: December 16, 2003, 08:56:52
Job time : 29.7392 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 26.2347 Seconds
(without alignments)
901.547 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPNVNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	143.5	5.0	2736	4	US-09-252-991A-30227	Sequence 30227, A
2	135	4.7	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
3	134.5	4.7	1638	4	US-09-071-035-258	Sequence 258, App
4	134.5	4.7	1638	4	US-09-071-035-262	Sequence 262, App
5	134.5	4.7	1638	4	US-09-071-035-266	Sequence 266, App
6	133	4.7	288	4	US-09-216-393B-341	Sequence 341, App
7	133	4.7	288	4	US-09-216-393B-344	Sequence 344, App
8	133	4.7	2314	4	US-09-268-347-49	Sequence 49, Appl
9	132.5	4.6	550	4	US-09-198-452A-613	Sequence 613, App
10	128	4.5	1070	4	US-09-107-532A-6735	Sequence 6735, Ap
11	127.5	4.5	558	4	US-09-071-035-268	Sequence 268, App
12	126	4.4	525	4	US-09-107-532A-5095	Sequence 5095, Ap
13	125.5	4.4	461	1	US-08-186-222-2	Sequence 2, Appli
14	125	4.4	1129	4	US-09-252-991A-29927	Sequence 29927, A
15	125	4.4	1702	3	US-08-296-791-5	Sequence 5, Appli
16	125	4.4	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
17	124.5	4.4	1529	2	US-08-728-470-10	Sequence 10, Appl
18	124.5	4.4	1529	3	US-08-719-641-10	Sequence 10, Appl
19	124.5	4.4	1600	2	US-08-617-697-10	Sequence 10, Appl
20	124	4.3	359	3	US-08-818-112-109	Sequence 109, App
21	124	4.3	359	4	US-08-818-111-104	Sequence 104, App
22	124	4.3	359	4	US-09-056-556-109	Sequence 109, App
23	124	4.3	359	4	US-09-072-596-104	Sequence 104, App
24	124	4.3	1619	4	US-09-328-352-7347	Sequence 7347, Ap
25	123	4.3	658	1	US-08-409-995-5	Sequence 5, Appli
26	123	4.3	658	3	US-08-685-467-5	Sequence 5, Appli
27	123	4.3	658	3	US-08-913-942-5	Sequence 5, Appli
28	123	4.3	1098	1	US-08-409-995-2	Sequence 2, Appli
29	123	4.3	1098	3	US-08-685-467-2	Sequence 2, Appli
30	123	4.3	1098	3	US-09-377-155-32	Sequence 32, Appl
31	123	4.3	1098	3	US-08-913-942-2	Sequence 2, Appli
32	123	4.3	1098	4	US-09-669-974-32	Sequence 32, Appl
33	123	4.3	1098	4	US-09-268-347-44	Sequence 44, Appl
34	123	4.3	1220	4	US-09-206-942-28	Sequence 28, Appl
35	123	4.3	1226	4	US-09-206-942-26	Sequence 26, Appl
36	123	4.3	1848	3	US-08-296-791-6	Sequence 6, Appli
37	123	4.3	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
38	123	4.3	2035	1	US-08-046-585-5	Sequence 5, Appli
39	123	4.3	2035	1	US-08-393-703-5	Sequence 5, Appli
40	123	4.3	2035	5	PCT-US93-11721-5	Sequence 5, Appli
41	122	4.3	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
42	121.5	4.3	1992	4	US-08-621-944A-3	Sequence 3, Appli
43	121.5	4.3	1992	4	US-08-945-567D-3	Sequence 3, Appli
44	121.5	4.3	2048	4	US-09-268-347-48	Sequence 48, Appl
45	121	4.2	833	4	US-09-252-991A-32338	Sequence 32338, A

ALIGNMENTS

RESULT 1
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match 5.0%; Score 143.5; DB 4; Length 2736;
Best Local Similarity 22.1%; Pred. No. 0.0037;
Matches 141; Conservative 56; Mismatches 236; Indels 205; Gaps 33;

```
Qy          2 PIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDN 61
              ||| :  : : | |  |  |||:        | ::|:|    |         |: | |
Db        723 PIGQVTADGSGNWSFTPTTPLPN--------GTVVNATATDAS-------GNTSAGSSVT 767

Qy         62 RASDVPGLPV-NPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120
                |  |  || ||    ::  ||:   |               ||||    |  ||
Db        768 VDSVAPATPVINP----SNGTTLSGTAEP-------------GSSV----TLTDGNGNP 805

Qy        121 VGQRN-------GVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGA---GHAMVTVAS 170
              :||              |  |:|       : || |       | |:|     |  |  |
Db        806 IGQVTADGSGNWSFTPSTPLADGTVVNATATDPAGN-----TSGQGSTTVDGVAPTTPTV 860

Qy        171 DITEARQRILELLEPKGTGESKGA-----GESKGVGELRESNSGAENTTETQTSTSTSSL 225
              :::            || |          |  : |:    ||  | | |
Db        861 NLSNGSSL-------SGTAEPGSTVILTDGNGNPIAEVTADGSG--NWTYT--------- 902

Qy        226 RSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPT-------------TTDPDAA 271
                |   :| |||   :   ||       ::| :  :|               | :| |
Db        903 ---PSTPIANGTVVNVVAQDAAGNSSPGASVTVDSQAPAAPVVNPSNGTTLSGTAEPGAT 959

Qy        272 ASATETATRDQLTKEAFQNPDNQKVNIDELGN--AIPSGVLKDDVVANIEEQAKAAGEEA 329
              | |          :   ||  | | |  ||     |   | :  | |           |
Db        960 VSLT----------DGNGNPIGQ-VTADGSGNWSFTPGTPLANGTVVNATASDPTGNTSA 1008
```

```
Qy        330 -KQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQ 388
                  ::: | |    :             : :|  |  |  | :  |  | |
Db       1009 PASTTVDSVAPAAPVVNPSNG-----VVISGTAEPGATVTLTDGSG------------N 1050

Qy        389 PVEQTTT------TTTTTTTTSARTVEN----KPANNTPAQGN--VD--TPGSEDTMESR 434
              |: | |         : | :|  :  || |        || ||  ||:  ||   :      |
Db       1051 PIGQVTADGSGNWSFTPSTPLADGTVVNATATDPAGNTGGQGSTTVDAIAPATPTVNLSN 1110

Qy        435 RSSMASTS---STFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRI-------W- 483
                ||:: |:   ||   |    | |                          :  :       |
Db       1111 GSSLSGTAEPGSTVILTDGNGNP-----------------------IAEVTADGSGNWT 1146

Qy        484 ---------GIQISVVYSTI---QHPPRDTT------------DNGARLLGNPSAGIQS 518
                        |  ::||          ||   |                 ||  :  |    ||
Db       1147 YTPSTPIANGTVVNVVAEDAAGNSSPPATVTVDSSAPPAPVINPSNGVVISGTAEAGATV 1206

Qy        519 TYARLALSGGLRHDMGGLTG-GSNSAVNTSNNPPAPGS 555
              |    |  :||  : :| :|  || :    |    ||| |:
Db       1207 T---LTDAGG--NPIGQVTADGSGNWSFTPGTPPANGT 1239
```

RESULT 2
US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

```
  Query Match             4.7%;  Score 135;  DB 4;  Length 2137;
  Best Local Similarity   17.9%;  Pred. No. 0.014;
  Matches   86;  Conservative   82;  Mismatches  235;  Indels   78;  Gaps   13;

Qy         36 INSTGPLGSRAL----FTPVRNSMADSGDNRASDVPGLPVNPMRLAASEIT-LNDGFEVL 90
              | ||   |::      :    ||| :||              || :  : |  | |:| :
Db        776 IESTDASGNKTTTKINYEVTRNSASDSTSTSI-------VNSVSTSISNSTSLSDSVKAS 828

Qy         91 HDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSVVLSD--QEYAR-------- 140
                        :|:  : :|      : :  :  :  ::     |   ||   | ||
Db        829 QSLSTSKSLSESLSASTSNSTSIQASESASTSKQLSESASTSTSDSASESARKSESTSKS 888

Qy        141 ---LQSIDPEGKDKFVFTGGRGGAGHAMV-----TVASDIT------------EARQRI 179
                 :|      |    :     :     |     |   |||                |:
Db        889 TSLSESTSTSVSDSASVSTSESASTSTSVSGSTSTSISDSTSTSTSDSASIKASESASTS 948

Qy        180 LELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVA 239
                 |   |    |  |  |  |  |  :    :::    :|  | |||:  ||         | :  :
Db        949 KLLSESVSTSTSDSASTSTSVSDSNSASTSLSKSTSTSVSDSTSTSTSDS----ASTSTS 1004

Qy        240 TGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLT-----------KEA 287
                  |:|  :  : : :  :     : |:|    |:  ||:|: :  :    |               :
Db       1005 ESESDSASTSLSESTSTSVSDSTSTSTSDSASMSASESESNSKSTSLSESTSTSLSGSTS 1064

Qy        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAA-GEEAKQQAIENNAQAQKK--- 343
                   |:   :   |   :   |   |  :     ::        |:  :  |        |::: ::
Db       1065 ASTSDSASTSTSESESDSTSTSLSESTSTSLSGSTSASTSDSASTSTSESDSTSESTSLS 1124

Qy        344 -------YDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTT- 395
                     |    |    |    |:      |.:   | |  :   ::: :           |:|
Db       1125 ESLSTSVSDSTSASTSESASTSTSESESNSASTSLSGSLSTSISDSTSTSTSDSASTSTS 1184

Qy        396 ------TTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTS 449
                    |:|: : ::: :: :  : :|     :   |   |: |  ||   |     |||::  |::
Db       1185 ESESDSTSTSLSESTSTSLSDSTSTSTSESASTSTSESDSTSES-TSLSESTSTSVSDST 1243

Qy        450 S 450
              |
Db       1244 S 1244
```

RESULT 3
US-09-071-035-258
; Sequence 258, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 258:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-258

```
  Query Match             4.7%;  Score 134.5;  DB 4;  Length 1638;
  Best Local Similarity   18.9%;  Pred. No. 0.01;
  Matches  139;  Conservative   79;  Mismatches  280;  Indels  239;  Gaps   30;

Qy          6 LGHNPNVNNSIPP----------APPLPSQTDGAGGRGQLINSTGPLGS---------- 44
                |  :|  | | |            |||.  :   |    :|   :   |  ||:
Db        867 LAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRATDGTV 926

Qy         45 ------------------RALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEITLND 85
                                :|       :|      |       |     :   |     :  |
Db        927 LGTATTGPTGQYTVTLASGKATAKQTVNVVA-KNDTGLESQPTTAMTPADVTTP--TIGD 983

Qy         86 -------GFEVLHDHGPLDTLN-RQIGSSVFRVETQEDGKHIAV----GQRNGVETSVVL 133
                     |:|:       |   |:  |       ::    |  :|   :   |      |  |  :|  |:
Db        984 ITGDSTTGYEITGTADPNTTIEVRNPDGTIIGTTTTDDQGNFTVDLPAGAANPGDTLTVV 1043

Qy        134 SDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILEL----------- 182
                            |||      |       :       |||   | | |      :   :
Db       1044 --------------GKD-----GDGNESQPTEVTVPEDATVAAPTVTTVTGTTATGYQVT 1084
```

```
Qy        183 --LEPK-------------GTGESKGAG------------ESKGVGELRESNSGAE-NTT 214
                ||                || : |||             :: :  : :  :| | | |
Db       1085 GTAEPNVTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEALTAIAKDAAGKESNPT 1144

Qy        215 ETQTSTSTSSLRSDPKLWLALGTVATG--------------------LIGLAATGIVQA 253
              :|      :  : | :    |:   |                    ::|:| ||
Db       1145 AFKTPADPDAPVATPTVDKITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTGTDGK 1204

Qy        254 LALTPEPDSPTTTD---------------PDAAASATE--TATRDQLTKEAFQNPD---- 292
              :| ||   :  :                 |  | :  :  | | | :|  : :  :
Db       1205 YTVTLEPGKASANETITVVAKNATGKESQPATATTPVDLATPTIDSITGNSSKGYEITGT 1264

Qy        293 -NQKVNID---------------ELGN---AIPSGVL----------KD----------- 312
               |  ||                | |     :|:||:           ||
Db       1265 AEPKTTIDVRDADGTIIAATTANETGQYTVTLPAGVVTPGETITIISKDGAGNESQPATA 1324

Qy        313 ----DVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGA 368
                  |||         |  | :|    :       |       |:|             ||
Db       1325 VIPADVVLAAPTITKVEGNKANGYTV--TGTADPNVTVQFYNSSEQLLA--------SGN 1374

Qy        369 LILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSE 428
              ||   | : | |  : : :  |  || |    | :|   ||: |   |  :   :
Db       1375 TTTGGTFSVHIAAGLATEKETL--TALTTDTQGNVSPKTTFMTPADIT---GEPEIKIAA 1429

Qy        429 DTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWG---I 485
              |:    ||:  ||   :       | ||  : : :| :    | :| :       |   |
Db       1430 PTV----SSVLGTSKAGYLIKGTAEPNRI--IQISNRLL----RSVIAVGATDAEGNFAI 1479

Qy        486 QISVVYSTIQHPPRDTTDNGARLLG------------NPSAGIQSTYARLALSGGLRHDM 533
              |::   :| |      | :||                || |  :|     :||   :
Db       1480 QLTAGQATAQQSLLATATDGAGHYSTATTFMTPADPTNPGGGNGNTGGNNGNTGGNTGNN 1539

Qy        534 GGLTGGSNSAVNTSNNP 550
              |   | : :  || :||
Db       1540 GATGGNNGNGSNTGSNP 1556
```

```
RESULT 4
US-09-071-035-262
; Sequence 262, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  262:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-262

  Query Match             4.7%;  Score 134.5;  DB 4;  Length 1638;
  Best Local Similarity   18.9%;  Pred. No. 0.01;
  Matches  139;  Conservative   79;  Mismatches  280;  Indels  239;  Gaps   30;

Qy          6 LGHNPNVNNSIPP----------APPLPSQTDGAGGRGQLINSTGPLGS----------- 44
              |  :|  | | |           |||   :  |   :|  :  |  ||:
Db        867 LAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRATDGTV 926

Qy         45 -------------------RALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEITLND 85
                               :|      | :|   |       |   : |  :     |: |
Db        927 LGTATTGPTGQYTVTLASGKATAKQTVNVVA-KNDTGLESQPTTAMTPADVTTP--TIGD 983

Qy         86 -------GFEVLHDHGPLDTLN-RQIGSSVFRVETQEDGKHIAV----GQRNGVETSVVL 133
                     |:|:      |  |:  |    ::    | :|  :  |     |  |  :|  |:
Db        984 ITGDSTTGYEITGTADPNTTIEVRNPDGTIIGTTTTDDQGNFTVDLPAGAANPGDTLTVV 1043

Qy        134 SDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILEL----------- 182
                          |||     |    :    |||  | | |    :. :
Db       1044 -------------GKD-----GDGNESQPTEVTVPEDATVAAPTVTTVTGTTATGYQVT 1084

Qy        183 --LEPK-------------GTGESKGAG------------ESKGVGELRESNSGAE-NTT 214
                ||                || : |||             :: :  : :  :| | | |
Db       1085 GTAEPNVTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEALTAIAKDAAGKESNPT 1144

Qy        215 ETQTSTSTSSLRSDPKLWLALGTVATG--------------------LIGLAATGIVQA 253
              :|      :  : | :    |:   |                    ::|:| ||
Db       1145 AFKTPADPDAPVATPTVDKITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTGTDGK 1204

Qy        254 LALTPEPDSPTTTD---------------PDAAASATE--TATRDQLTKEAFQNPD---- 292
              :| ||   :  :                 |  | :  :  | | | :|  : :  :
Db       1205 YTVTLEPGKASANETITVVAKNATGKESQPATATTPVDLATPTIDSITGNSSKGYEITGT 1264

Qy        293 -NQKVNID---------------ELGN---AIPSGVL----------KD----------- 312
               |  ||                | |     :|:||:           ||
Db       1265 AEPKTTIDVRDADGTIIAATTANETGQYTVTLPAGVVTPGETITIISKDGAGNESQPATA 1324

Qy        313 ----DVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGA 368
                  |||         |  | :|    :       |       |:|             ||
Db       1325 VIPADVVLAAPTITKVEGNKANGYTV--TGTADPNVTVQFYNSSEQLLA--------SGN 1374

Qy        369 LILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSE 428
              ||   | : | |  : : :  |  || |    | :|   ||: |   |  :   :
Db       1375 TTTGGTFSVHIAAGLATEKETL--TALTTDTQGNVSPKTTFMTPADIT---GEPEIKIAA 1429

Qy        429 DTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWG---I 485
              |:    ||:  ||   :       | ||  : : :| :    | :| :       |   |
Db       1430 PTV----SSVLGTSKAGYLIKGTAEPNRI--IQISNRLL----RSVIAVGATDAEGNFAI 1479

Qy        486 QISVVYSTIQHPPRDTTDNGARLLG------------NPSAGIQSTYARLALSGGLRHDM 533
              |::   :| |      | :||                || |  :|     :||   :
Db       1480 QLTAGQATAQQSLLATATDGAGHYSTATTFMTPADPTNPGGGNGNTGGNNGNTGGNTGNN 1539

Qy        534 GGLTGGSNSAVNTSNNP 550
              |   | : :  || :||
Db       1540 GATGGNNGNGSNTGSNP 1556
```

```
RESULT 5
US-09-071-035-266
; Sequence 266, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  266:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-266

  Query Match             4.7%;  Score 134.5;  DB 4;  Length 1638;
  Best Local Similarity   18.9%;  Pred. No. 0.01;
  Matches  139;  Conservative   79;  Mismatches  280;  Indels  239;  Gaps   30;

Qy          6 LGHNPNVNNSIPP----------APPLPSQTDGAGGRGQLINSTGPLGS---------- 44
              |  :|  | | |           |||   :  |   :|  :  |  ||:
Db        867 LAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRATDGTV 926

Qy         45 ------------------RALFTPVRNSMADSGDNRASDVPGLPVNPMRLAASEITLND 85
                              :|     | :|    |       |   : |  :     |: |
Db        927 LGTATTGPTGQYTVTLASGKATAKQTVNVVA-KNDTGLESQPTTAMTPADVTTP--TIGD 983

Qy         86 -------GFEVLHDHGPLDTLN-RQIGSSVFRVETQEDGKHIAV----GQRNGVETSVVL 133
                  |:|:     |  |:  |    ::    | :|  :  |    |  |  :|  |:
Db        984 ITGDSTTGYEITGTADPNTTIEVRNPDGTIIGTTTTDDQGNFTVDLPAGAANPGDTLTVV 1043

Qy        134 SDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILEL---------- 182
                        |||      |    :    |||  | | |   :  :
Db       1044 -------------GKD-----GDGNESQPTEVTVPEDATVAAPTVTTVTGTTATGYQVT 1084

Qy        183 --LEPK------------GTGESKGAG-----------ESKGVGELRESNSGAE-NTT 214
                 ||                || : |||             :: :  :  :  :| | | |
Db       1085 GTAEPNVTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEALTAIAKDAAGKESNPT 1144

Qy        215 ETQTSTSTSSLRSDPKLWLALGTVATG--------------------LIGLAATGIVQA 253
                :|      :  : | :    |:   |                   ::|:| ||
Db       1145 AFKTPADPDAPVATPTVDKITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTGTDGK 1204

Qy        254 LALTPEPDSPTTTD--------------PDAAASATE--TATRDQLTKEAFQNPD---- 292
                :| ||   :  :              |  | :  :  | | | :|  : :  :
Db       1205 YTVTLEPGKASANETITVVAKNATGKESQPATATTPVDLATPTIDSITGNSSKGYEITGT 1264

Qy        293 -NQKVNID---------------ELGN---AIPSGVL----------KD---------- 312
                 |  ||               | |    :|:||:          ||
Db       1265 AEPKTTIDVRDADGTIIAATTANETGQYTVTLPAGVVTPGETITIISKDGAGNESQPATA 1324

Qy        313 ----DVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGA 368
                    |||    .  |  | :|   :     |      |     |:|          ||
Db       1325 VIPADVVLAAPTITKVEGNKANGYTV--TGTADPNVTVQFYNSSEQLLA--------SGN 1374

Qy        369 LILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSE 428
                 ||   | : | |  : : :  |  || |    | :|    ||: |   |  :   :
Db       1375 TTTGGTFSVHIAAGLATEKETL--TALTTDTQGNVSPKTTFMTPADIT---GEPEIKIAA 1429

Qy        429 DTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWG---I 485
                |:    ||:  ||   :       | ||  : : :| :    | :| :      |   |
Db       1430 PTV----SSVLGTSKAGYLIKGTAEPNRI--IQISNRLL----RSVIAVGATDAEGNFAI 1479

Qy        486 QISVVYSTIQHPPRDTTDNGARLLG-----------NPSAGIQSTYARLALSGGLRHDM 533
                |::   :| |      |  :||                 ||  |  :|      :||  :
Db       1480 QLTAGQATAQQSLLATATDGAGHYSTATTFMTPADPTNPGGGNGNTGGNNGNTGGNTGNN 1539

Qy        534 GGLTGGSNSAVNTSNNP 550
                |   | : :  || :||
Db       1540 GATGGNNGNGSNTGSNP 1556


RESULT 6
US-09-216-393B-341
; Sequence 341, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THERE
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 341
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393B-341

  Query Match             4.7%;  Score 133;  DB 4;  Length 288;
  Best Local Similarity   18.7%;  Pred. No. 0.00093;
  Matches   72;  Conservative   23;  Mismatches  133;  Indels  158;  Gaps    6;

Qy         94 GPLDTLNRQIGSSVFRVETQEDGKHIAVGQ--------------------RNGVETSVV 132
               ||:  | |    ||  |||      :|:                        :  ||
Db         18 GPVSQLARASDDSVDSVETARQHMELAIEADEEMHEAYDPLLEFVETFREIKKAVEEDAA 77

Qy        133 LSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLE------PK 186
                ||     |:  |                    :|::  | || |   :||
Db         78 LSTDAIDRVSQFD------------------LVSLLDVIREAAQAKFDLLGRLITDIAS 118

Qy        187 GTGESKGA--GESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATGLIG 244
               | ||   |  ||        | |  | : || | :|||||:            |  :
Db        119 GIGEGAMALMGEEAAFIRPRRSKRGKKTTTTSSSTSTSTT----------TTTSTTTT 167

Qy        245 LAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNA 304
                  |        ||   : |||      : | | |    |       |
Db        168 TTTTTTTTTTTTPTTTTTTTTTTPTTTTTTTTTTPTTTTTTTTTTPTT---------- 216

Qy        305 IPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYG 364

Db        217 ------------------------------------------------------------ 216

Qy        365 LSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDT 424
                                                 ||||||  |||||  |         ||      |
Db        217 ---------------------------TTTTTTPTTTTSTTTTTTTTTTTPTTTTTTT 248

Qy        425 PGSEDTMESRRSSMASTSSTFFDTSS 450
                :  |  :  ::  ||::|   |::
Db        249 EPTTTTTTTEPTTTSTTTTTTTTTT 274


RESULT 7
```

```
US-09-216-393B-344
; Sequence 344, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 344
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393B-344

  Query Match             4.7%;  Score 133;  DB 4;  Length 288;
  Best Local Similarity   18.7%;  Pred. No. 0.00093;
  Matches   72;  Conservative   23;  Mismatches  133;  Indels  158;  Gaps    6;

Qy         94 GPLDTLNRQIGSSVFRVETQEDGKHIAVGQ---------------------RNGVETSVV 132
              ||:  | |    ||  |||      :|:                       :  ||
Db         18 GPVSQLARASDDSVDSVETARQHMELAIEADEEMHEAYDPLLEFVETFREIKKAVEEDAA 77

Qy        133 LSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLE------PK 186
              ||     |:   |                    :|::   | || |   :||
Db         78 LSTDAIDRVSQFD------------------LVSLLDVIREAAQAKFDLLGRLITDIAS 118

Qy        187 GTGESKGA--GESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATGLIG 244
              | ||   |  ||       | |  | : || | :|||||:             |  :
Db        119 GIGEGAMALMGEEAAFIRPRRSKRGKKTTTTTSSSTSTSTT-----------TTTSTTTT 167

Qy        245 LAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNA 304
                |       ||   : |||      : | | |    |       |
Db        168 TTTTTTTTTTTTPTTTTTTTTTPTTTTTTTTTPTTTTTTTTTPTT----------- 216

Qy        305 IPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGAGYG 364

Db        217 ------------------------------------------------------------ 216

Qy        365 LSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDT 424
                                      |||||| |||||  |         ||       |
Db        217 --------------------------TTTTTTPTTTTSTTTTTTTTTTTTPTTTTTTT 248

Qy        425 PGSEDTMESRRSSMASTSSTFFDTSS 450
                :  |  :  ::  ||::|   |::
Db        249 EPTTTTTTTEPTTTSTTTTTTTTT 274


RESULT 8
US-09-268-347-49
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 2314
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-49

  Query Match             4.7%;  Score 133;  DB 4;  Length 2314;
  Best Local Similarity   21.3%;  Pred. No. 0.024;
  Matches  102;  Conservative   78;  Mismatches  191;  Indels  108;  Gaps   23;

Qy         26 TDGAGGRGQLINSTGPLGSRALFTPVRNSM-------------ADSGDNRASDVPGLPV 71
              |   |   | ||     |: ||    |  :                  :: :|  : ||
Db        245 TIAIGSNAQAIN----YGALALGADTRVDLDYGIALGYGSQILNNNNNNNNKAYVP--EG 298

Qy         72 NPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSV 131
              |   : :|: | |  |                  ||||  :       | | ||   | | :
Db        299 NGSNIKSSKATGNGLF-------------SIGSSTIK------RKIINVGA--GYEDTD 336

Qy        132 VLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKGTGES 191
               ::    |:|::::   | :  | |   | |                 :::: | |  || ||:
Db        337 AVN---VAQLKAVENLAKRQITFKGDDNGTG------------VKKKLGETLTIKG-GET 380

Qy        192 KG--AGESKGVGELRESNSG-----AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIG 244
              :     ::  :| : ::|:|     |:| :  :|  || :|  :  |:     || :|
Db        381 QADKLTDNNNIGVVTDNNTGLKVKLAKNLSGLET-VSTKNLTASEKV-----TVGSG--- 431

Qy        245 LAATGIVQALALTPEPDSPTTTD-----------PDAAASATETATRDQLTKEAFQNP-- 291
                |  :|:  ||  | :  :||             | : :| |  ||    |  | |
Db        432 -NNTAELQSGGLTFTPTTNASTDKTVYGTDGLKFTDNSNTALEDTTRITKDKIGFSNKAG 490

Qy        292 --DNQKVNID----ELGNAI--PSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKK 343
                |  |  :|    ::||:     |:  :: :    :| :   :  |   :  |     |
Db        491 TVDENKPYLDKDKLKVGNSTLNNGGLTVNNTIGGSNKQIQVGADGIKFADVNVNVSNAAK 550

Qy        344 YDEQQAKRQE------ELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQP---VEQTT 394
              :   :  :|       : ||   : | |    :  ||: :   : ::  :|    |:  |
Db        551 FGTTRITEEEIGFADADGKVDKKSPY-LDKKQLQVGGVKITKDSGINAGDQKISNVKDAT 609

Qy        395 TTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGG 453
                |   |   : |:       :   |  | | |: |   |  :| :| |:| :  |
Db        610 DDTDAVTYKQLKQVQQDADGALQSFSIRDEKGQEFTI-SNLYSNGNTPNT-FETITFAG 666


RESULT 9
US-09-198-452A-613
; Sequence 613, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 613
;   LENGTH: 550
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-09-198-452A-613

  Query Match             4.6%;  Score 132.5;  DB 4;  Length 550;
  Best Local Similarity   21.1%;  Pred. No. 0.0028;
  Matches  104;  Conservative   57;  Mismatches  166;  Indels  167;  Gaps   23;

Qy          5 NLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGP----------------------L 42
              ::|    :|    |||      :  ||     : |||                       |
Db        120 DVGGLKGINWKTTPAPDFSFINETPGGGAHSTSHTGPGTPVGATVVPNVNVNLGGIKVDL 179

Qy         43 GSRAL-------------FTPVRNSMADSGDNRAS-----------------DVPGL--- 69
              |   |             | : :: : | |:: |                   | ||
Db        180 GGINLGGITTNVTTEEGGGTNITSTKSTSTDDKVSITSTGSQSTIEEDTIQFDDPGQGED 239

Qy         70 ------------PVNPMRLAASEI------TLNDGFEVLHDHGPLDTLNRQIGSSVFRVE 111
                         |  |  |::| :      :||   : :  |  |:|     |:||   :
Db        240 DNAIPGTNTPPPPGPPPNLSSSRLLTISNASLNQVLQNVRQH--LNTAYDSNGNSV--SD 295

Qy        112 TQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGG-RGGAGHAMVTVAS 170
               :|   :   |||    |:  :        || |:     |||   | ||
Db        296 LNQDLGQVVKNSENGVNFPTVILPKTTG---DTDPSGQ----ATGGVTEGGGH------- 341
```

```
Qy        171 DITEARQRILELLEPKGTGESKGAGE------SKGVGELRESNSGAEN----------T 213
Db        342 -IRNIIQR-----NTQSTGQSEGATPTPQPTIAKIVTSLRKANVSSSSVLPQPQVATTIT 395

Qy        214 TETQT-STSTSSLRSDPKLWLALGT-VATGLIGLAATGIVQALALTPEPDSPTTTDPDAA 271
Db        396 PQARTASTSTTSIGTGTESTSTTSTGTGTGSVSTQSTGV----------GTPTTTTRSTG 445

Qy        272 ASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQ 331
Db        446 TSATTTT-----SSASTQTPQ----------APLPSGTRH---VATISLVRNAAG----- 482

Qy        332 QAIENNAQAQKKYDEQQAKRQEELKV-SSGAGYGLSGALI------LGGGIGVAVTAALH 384
Db        483 ----------RSIVLQQGGRSQSFPIPPSGTGTQNMGAQLWAAASQVASTLGQVVNQAAT 532

Qy        385 RKNQPVEQTTTTTT 398
Db        533 AGSQPSSRRSSPTS 546


RESULT 10
US-09-107-532A-6735
; Sequence 6735, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 6735:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1070 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...1070
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6735:
US-09-107-532A-6735

  Query Match             4.5%;  Score 128;  DB 4;  Length 1070;
  Best Local Similarity   20.7%;  Pred. No. 0.02;
  Matches  104;  Conservative   69;  Mismatches  181;  Indels  148;  Gaps   24;

Qy         25 QTDGAGGRGQLINSTGP-------LGSRALFTPVRNSMADSGDNRASDVPGLPVNPMRLA 77
Db        329 RADGVSDEGEAVD-TLPGNVMFEVSGSGSEFS-VNTKINITNDNNRG------IIQMRGS 380

Qy         78 ASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQR------------- 124
Db        381 GSKVTVSNGGKI--------TINTAT-TGGFRLQGDSSQIHVSSGGQISVKMAGDDDQPG 431

Qy        125 -NGVE-------------TSVVLSDQEYARLQSIDPE-GKDKFVFTGG------RGGAGH 163
Db        432 NNGMRFVGKGLSLIVEGAGSVIDIDKQSGRSSAVRFENGTQKLTVTNGGSLRVRNTGDGK 491

Qy        164 AMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELR----ESNSGAENTTETQTS 219
Db        492 SYVNRLNQGNQA----IQFYDASGFLESPGSADFTVTGEKSNIDIRADSGA--TVDT-TG 544

Qy        220 TSTSSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATET-- 277
Db        545 DIDLNFTAGEKTYL----VMVGKTGEDKTGIISGDVLNVDLESPTYFDFQNKRIGSETNG 600

Qy        278 ----------------------------ATRDQLTKEA--FQNPDNQKVNIDELGNAIPS 307
Db        601 IGGWVFQGNGDSSLKIAHSEIALWRKTGATFDNFEKDPDYFSDMTDLSINGTDLGNFVSS 660

Qy        308 GVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQ----EELKVSSGAGY 363
Db        661 ------------PSAGLTNEFAQTGA------GMKNYNRISANNQLPMIESLRVPTNADK 702

Qy        364 GLSGALILGGGI----------GVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPA 413
Db        703 KIYGHVVVPEGVEAKPRDAWTGEVEVTLRI----TYADTTKSSVEITALTQGQTDENEGA 758

Qy        414 NNTP----AQG---NVDTPGSE 428
Db        759 GYNPWGEGEQGGLFEVQVPNDE 780


RESULT 11
US-09-071-035-268
; Sequence 268, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
```

```
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  268:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  558 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-268

  Query Match             4.5%;  Score 127.5;  DB 4;  Length 558;
  Best Local Similarity   20.3%;  Pred. No. 0.0081;
  Matches  116;  Conservative   61;  Mismatches  261;  Indels  133;  Gaps   20;

Qy         10 PNVNNSIPPAPPL---PSQTDGAGGRGQLINSTGPLG----SRALFTPVRNSMADSGDNR 62
Db         44 PNVTIEIHNEAGLVIATGTTDGAGA----FTITLPTGTATANEALTAIAKDAAGKESNPT 99

Qy         63 ASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVG 122
Db        100 AFKTPADPDAPV-------------------ATPTVDKITGSTT-------NGYQVVGA 132

Qy        123 QRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQR---- 178
Db        133 AEVGTTVEVRDADGTVLGMATTGTDGKYTVTLEPGKASANETITVVAKNATGKESQPATA 192

Qy        179 --ILELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALG 236
Db        193 TTPVDLATP--TIDSITGNSSKGY-----------EITGTAEPKTTIDVRD--------- 230

Qy        237 TVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKV 296
Db        231 ----------ADGTIIAATTANETGQYTVTLPAGVVTPGETIT--------------II 265

Qy        297 NIDELGNAI--PSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEE 354
Db        266 SKDGAGNESQPATAVIPADVVLAAPTITKVEGNKANGYTV--TGTADPNVTVQFYNSSEQ 323

Qy        355 LKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVENKPAN 414
Db        324 LLA--------SGNTTTGGTFSVHIAAGLATEKETL--TALTTDTQGNVSPKTTFMTPAD 373

Qy        415 NTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLL 474
Db        374 IT---GEPEIKIAAPTV----SSVLGTSKAGYLIKGTAEPNRI--IQISNRLL----RSV 420

Qy        475 ILIRLFRIWG---IQISVVYSTIQHPPRDTTDNGARLLG------------NPSAGIQST 519
Db        421 IAVGATDAEGNFAIQLTAGQATAQQSLLATATDGAGHYSTATTFMTPADPTNPGGGNGNT 480

Qy        520 YARLALSGGLRHDMGGLTGGSNSAVNTSNNP 550
Db        481 GGNNGNTGGNTGNNGATGGNNGNGSNTGSNP 511


RESULT 12
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 5095:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 525 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...525
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095

  Query Match             4.4%;  Score 126;  DB 4;  Length 525;
  Best Local Similarity   21.8%;  Pred. No. 0.01;
  Matches   77;  Conservative   66;  Mismatches  152;  Indels   58;  Gaps   13;

Qy        128 ETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEPKG 187
Db         32 DSQIQQQDQKIADLKNQQADAQSQ------------IDALESQVSEINTQAQDLLAKQD 78

Qy        188 TGESKGAGESKGVGELRESNSGAENT-----TETQTSTSTSS-----LRSDPKLWLALGT 237
Db         79 TLRQESAQLVKDIADLQERIEKREDTIQKQAREAQVSNTSSNYIDAVLNAD-SLADAIGR 137

Qy        238 VATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNP---DNQ 294
Db        138 VQ------AMTTMVKA-----NNDLMEQQKQDKKAVEDKKAENDAKLKELAENQAALESQ 186

Qy        295 KVNIDELGNAIPSGVLK----------DDVVANIEEQ-AKAAGEEAK---QQAIENNAQA 340
Db        187 KG--DLLSKQADLNVLKTSLAAEQATAEDKKADLNRQKAEAEAEQARIREQQRLAEQARQ 244

Qy        341 QKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTT 400
Db        245 QAAQEKAEKEAREQAEAEAQATQASSTAQSSATEESSATQSSMTEESSSATQSSATEEST 304

Qy        401 TTTSARTVENKPANNTPAQGNVDTPGSEDTMESR---RSSMASTSSTFFDTSS 450
Db        305 TPESS-TEESTAPESSATEESTTAPESSATEESTTVPESSATEESTTVPESST 356


RESULT 13
US-08-186-222-2
; Sequence 2, Application US/08186222
; Patent No. 5559007
;  GENERAL INFORMATION:
;    APPLICANT:  Suri, Bruno
;    APPLICANT:  Schmitz, Albert
;    TITLE OF INVENTION:  Bacterial Vectors
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
```

```
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/186,222
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,205
;      FILING DATE:  19-MAR-1991
;      APPLICATION NUMBER:  GB 9006400.7
;      FILING DATE:  22-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Villamizar, JoAnn
;      REGISTRATION NUMBER:  30,598
;      REFERENCE/DOCKET NUMBER:  4-17994/A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914)785-7121
;      TELEFAX:  (914)347-5769
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  461 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-186-222-2

  Query Match             4.4%;  Score 125.5;  DB 1;  Length 461;
  Best Local Similarity   19.0%;  Pred. No. 0.0091;
  Matches   84;  Conservative   70;  Mismatches  160;  Indels  127;  Gaps   16;

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTST--SSLRSDPK-LWLALGT 237
              :: :   |  |  : :::   ::    |  ::  : ||||    : : |: | |   : |
Db         32 DIAKQDATISSAQSAKAQAQAQVDSLQSKVDSLQQKQTSTKAQIAKIESERKALNAQIAT 91

Qy        238 VATGLIGLAATGIVQALALTPEPDSPTTTDPDA-------------AASATETATRDQL 283
              :   :     |  ::| | : : :|  |   ||             | || ::   |:
Db         92 LNESI--KERTKTLEAQARSAQVNSSATNYMDAVVNSKSLTDVIQKVTAIATVSSANKQM 149

Qy        284 -------TKEAFQNPDNQKVNIDE---LGNAIPS---------GVLKDDVVANIEEQAKA 324
                    ||  |  :  | | ::   |  :: |           ||   || :  ||
Db        150 LEQQEKEQKELSQKSETVKKNYNQFVSLSQSLDSQAQELTSQQAELK---VATLNYQATI 206

Qy        325 AGEEAKQQAIENNAQAQKKYDEQQAKRQE--ELKVSSGAGYGLSGALILGGGIGVAVTAA 382
              |  : |:||: :   | :|  :: ||:|   | :    |   :          :  | ::|
Db        207 ATAQDKKQALLDEKAAAEKAAQEAAKKQAAYEAQQKEAAQAQAASTAATAKAVEAATSSA 266

Qy        383 LHRKNQPVEQTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRRSSMASTS 442
                :|  : :|:|  ||:  ||     |  :|:: :  :     |  :  |  |:    :
Db        267 SASSSQAPQVSTSTDNTTSNASASNSSNSSSNSSSSSSS----SSSSSSSSSNSNAGGNT 322

Qy        443 STFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTT 502
              ::   | : ||                                              ||
Db        323 NSGTSTGNTGG-----------------------------------------------TT 335

Qy        503 DNGARL----LGNPSA-----------GIQSTYARLALSG-----------GLRHDMG 534
                |: :    :||| |           |  | |  : |             |: | :|
Db        336 TGGSGINSSPIGNPYAVGGCTDYVWQYFAAQGIYIRNIMPGNGGQWASNGPAQGVLHVVG 395

Qy        535 ---GLTGGSNSA--VNTSNNP 550
                |:   | ||  |  :|:|
Db        396 AAPGVIASSFSADFVGYANSP 416


RESULT 14
US-09-252-991A-29927
; Sequence 29927, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 29927
;   LENGTH: 1129
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29927

  Query Match             4.4%;  Score 125;  DB 4;  Length 1129;
  Best Local Similarity   21.4%;  Pred. No. 0.041;
  Matches   92;  Conservative   60;  Mismatches  168;  Indels  110;  Gaps   18;

Qy         44 SRALFTPVRNSMADSGDNRASDVPGLPVNPMRLAAS-----------EITLN------- 84
              |:| |:   :::      |  | : |       |:|||           ::| |
Db         89 SQAQFSQSMDAVYSVAQNAQSSLEGTAELYQRIAASTGDLGVNQQQVVQVTQNISKAMSA 148

Qy         85 DGFEVLHDHGPLDTLNRQIGSSVFR--------------VETQEDGKHIAVGQ-----RN 125
                |       | |  | :   | | |                 :   :|  :|||
Db        149 SGVSAAAAEGALVQLGQAFASGVLRGQELNSVLEQAPGLAQAVANGLGVAVGDLRKLGEQ 208

Qy        126 GVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILELLEP 185
              |  ||  : :   :: ::||     |:|          |  |:|          ::||
Db        209 GKLTSKQVFEAILSQTRAID----DQF---------ARAQTTIAGAF-----QVLENSAT 250

Qy        186 KGTGE-SKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVATGLIG 244
              :  |       | ||    |   | | : ::|  |:         :  |:||:|  |  |:|
Db        251 RAIGSLDSTLGVSKAFTEAMVSLSKSLDSTNVQSFVQV----LNTGLYLAIGRTAGALVG 306

Qy        245 LAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNA 304
              |  |   : :| |       |   : |:| |       :  ||          :|    |
Db        307 ATAAKIAD-VKVTQE-----QTYAASVAAAGEVRRAQAVKAEAV-------AELDRARQA 353

Qy        305 IPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAK-------------R 351
              : |   :  | |: | |   | | :: :|:: :  |::: : |: |             |
Db        354 VASA--RAQVAADRERQ---ASEISRLRAVQASLVAERELEGQRLKAQITEIGRQQSVAR 408

Qy        352 QEELKVS--------SGAGYGLSGALILGGGIGVAVTAALHRKNQPVE-------QTTTT 396
              ||:::           |   |:   : |     ||||| ::    |         | |
Db        409 MAELRLAETAIIKQLQAAEAQLTATTVAG---SQAVTAALAQRVSATEALSAANLQLTAA 465

Qy        397 TTTTTTTSAR 406
                | :|    |
Db        466 QTASTAAMGR 475


RESULT 15
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
```

```
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1702 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-5

  Query Match             4.4%;  Score 125;  DB 3;  Length 1702;
  Best Local Similarity   17.8%;  Pred. No. 0.078;
  Matches   87;  Conservative   62;  Mismatches  180;  Indels  160;  Gaps   17;

Qy         10 PNVVNNSI--------PPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDN 61
              |: |  |         ||||  ||:|                  |   ||  :|
Db       1017 PSNNEEIARVETPVPPPAPATPSETTE------------------TVAENSKQES--- 1053

Qy         62 RASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAV 121
                        |      |:| |  :| ||  :  |    | |        |  : |
Db       1054 -------KTVEKNEQDATETTAQNG-EVAEEAKPSVKANTQTN------EVAQSGSETEE 1099

Qy        122 GQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRILE 181
               |   :: :  :  :| |:::  :    :|                    :| || |   |
Db       1100 TQTTEIKETAKVEKEEKAKVEKEEKAKVEK-----------------DEIQEAPQMASE 1141

Qy        182 LLEPKGTGESKGAGESKGVGELR-ESNSGAENTT--------------ETQTSTSTSSL 225
                 :    |         | | : ::     ::||               |||  |  |:
Db       1142 TSPKQAKPAPKEVSTDTKVEETQVQAQPQTQSTTVAAAEATSPNSKPAEETQPSEKTN-- 1199

Qy        226 RSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTK 285
               ::|                  | :|          | ||  :  |     :|    |:
Db       1200 -AEP----------------VTPVVSKNQTENTTDQPTEREKTAKVETEKTQEPPQVAS 1241

Qy        286 EAFQNPDNQKV----------NIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIE 335
               :|    :  :           |:  : ||       ::| | :: |  |     || | |
Db       1242 QASPKQEQSETVQPQAVLESENVPTVNNA-------EEVQAQLQTQTSAT-VSTKQPAPE 1293

Qy        336 NNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTT 395
               |:                                :  |   |:|    : ::|  :|
Db       1294 NS--------------------------------INTGSATAITETAEKSDKPQTETAA 1320

Qy        396 TTTTTTTTSARTV-ENKPANNTPA-------QGNVDTPGSEDTMESRRSSMASTSSTFFD 447
               :|   :   | || :|  |||: :       : ::  |  ::|     :: ::  :|  |
Db       1321 STEDASQHKANTVADNSVANNSESSEPKSRRRRSISQP--QETSAEETTAASTDETTIAD 1378

Qy        448 TSSIGGPCR 456
                |    | |
Db       1379 NSKRSKPNR 1387


Search completed: December 16, 2003, 08:57:56
Job time : 29.2347 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:55:50 ; Search time 27.2437 Seconds
(without alignments)
3816.109 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPNVNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	148.5	5.2	596	11	US-09-946-374-243	Sequence 243, App
2	148.5	5.2	596	12	US-10-015-387A-243	Sequence 243, App
3	148.5	5.2	596	12	US-10-063-735-100	Sequence 100, App
4	148.5	5.2	596	12	US-10-006-130A-243	Sequence 243, App
5	148.5	5.2	596	12	US-10-199-672-310	Sequence 310, App
6	148.5	5.2	596	12	US-10-006-172A-243	Sequence 243, App
7	148.5	5.2	596	12	US-10-187-749-310	Sequence 310, App
8	148.5	5.2	596	12	US-10-194-457-310	Sequence 310, App
9	148.5	5.2	596	12	US-10-184-642-310	Sequence 310, App
10	148.5	5.2	596	12	US-10-196-747-310	Sequence 310, App
11	148.5	5.2	596	12	US-10-015-392A-243	Sequence 243, App
12	148.5	5.2	596	12	US-10-017-253A-243	Sequence 243, App
13	148.5	5.2	596	12	US-10-173-689-310	Sequence 310, App
14	148.5	5.2	596	12	US-10-173-690-310	Sequence 310, App
15	148.5	5.2	596	12	US-10-173-691-310	Sequence 310, App
16	148.5	5.2	596	12	US-10-173-692-310	Sequence 310, App
17	148.5	5.2	596	12	US-10-173-694-310	Sequence 310, App
18	148.5	5.2	596	12	US-10-173-698-310	Sequence 310, App
19	148.5	5.2	596	12	US-10-173-699-310	Sequence 310, App
20	148.5	5.2	596	12	US-10-173-707-310	Sequence 310, App
21	148.5	5.2	596	12	US-10-174-569-310	Sequence 310, App
22	148.5	5.2	596	12	US-10-174-583-310	Sequence 310, App
23	148.5	5.2	596	12	US-10-174-587-310	Sequence 310, App
24	148.5	5.2	596	12	US-10-174-589-310	Sequence 310, App
25	148.5	5.2	596	12	US-10-174-591-310	Sequence 310, App
26	148.5	5.2	596	12	US-10-175-736-310	Sequence 310, App
27	148.5	5.2	596	12	US-10-175-742-310	Sequence 310, App
28	148.5	5.2	596	12	US-10-175-744-310	Sequence 310, App
29	148.5	5.2	596	12	US-10-175-745-310	Sequence 310, App
30	148.5	5.2	596	12	US-10-175-748-310	Sequence 310, App
31	148.5	5.2	596	12	US-10-175-751-310	Sequence 310, App
32	148.5	5.2	596	12	US-10-175-754-310	Sequence 310, App
33	148.5	5.2	596	12	US-10-176-480-310	Sequence 310, App
34	148.5	5.2	596	12	US-10-176-489-310	Sequence 310, App
35	148.5	5.2	596	12	US-10-176-754-310	Sequence 310, App
36	148.5	5.2	596	12	US-10-176-755-310	Sequence 310, App
37	148.5	5.2	596	12	US-10-176-759-310	Sequence 310, App
38	148.5	5.2	596	12	US-10-176-920-310	Sequence 310, App
39	148.5	5.2	596	12	US-10-176-922-310	Sequence 310, App
40	148.5	5.2	596	12	US-10-176-924-310	Sequence 310, App
41	148.5	5.2	596	12	US-10-176-984-310	Sequence 310, App
42	148.5	5.2	596	12	US-10-179-508-310	Sequence 310, App
43	148.5	5.2	596	12	US-10-179-512-310	Sequence 310, App
44	148.5	5.2	596	12	US-10-179-515-310	Sequence 310, App
45	148.5	5.2	596	12	US-10-017-306A-243	Sequence 243, App

ALIGNMENTS

RESULT 1
US-09-946-374-243
; Sequence 243, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C1
; CURRENT APPLICATION NUMBER: US/09/946,374
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750

; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102571
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102684
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102687
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102965
; PRIOR FILING DATE: 1998-10-02
; PRIOR APPLICATION NUMBER: 60/103258
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103314
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103315
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257

```
;  PRIOR FILING DATE: 1998-10-14
;  PRIOR APPLICATION NUMBER: 60/104987
;  PRIOR FILING DATE: 1998-10-20
;  PRIOR APPLICATION NUMBER: 60/105000
;  PRIOR FILING DATE: 1998-10-20
;  PRIOR APPLICATION NUMBER: 60/105002
;  PRIOR FILING DATE: 1998-10-20
;  PRIOR APPLICATION NUMBER: 60/105104
;  PRIOR FILING DATE: 1998-10-21
;  PRIOR APPLICATION NUMBER: 60/105169
;  PRIOR FILING DATE: 1998-10-22
;  PRIOR APPLICATION NUMBER: 60/105266
;  PRIOR FILING DATE: 1998-10-22
;  PRIOR APPLICATION NUMBER: 60/105693
;  PRIOR FILING DATE: 1998-10-26
;  PRIOR APPLICATION NUMBER: 60/105694
;  PRIOR FILING DATE: 1998-10-26
;  PRIOR APPLICATION NUMBER: 60/105807

  Query Match           5.2%;  Score 148.5;  DB 11;  Length 596;
  Best Local Similarity  17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|              | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:  :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |     |    |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :   :   ::|||   | :::||::|           : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |    ::|                    |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||      | :     | | |  :     :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
                :: | :  | |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                ||  :    || :||   |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 2
US-10-015-387A-243
; Sequence 243, Application US/10015387A
; Publication No. US20030135034A1
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan l.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2830P1C54
;  CURRENT APPLICATION NUMBER: US/10/015,387A
;  CURRENT FILING DATE:  2001-12-12
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-015-387A-243

  Query Match           5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity  17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|              | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:  :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |     |    |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :   :   ::|||   | :::||::|           : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |    ::|                    |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||      | :     | | |  :     :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
                :: | :  | |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                ||  :    || :||   |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 3
US-10-063-735-100
; Sequence 100, Application US/10063735
; Publication No. US20030138882A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,735
;  CURRENT FILING DATE:  2002-05-08
;      Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-735-100

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::| :   |  : |:|              | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |    |    |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|          : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |   ::|                       |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||    | :    | | |  :     :  :   | : ::| ::  | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                              ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 4
US-10-006-130A-243
; Sequence 243, Application US/10006130A
; Publication No. US20030148375A1
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan l.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2830P1C7
;  CURRENT APPLICATION NUMBER: US/10/006,130A
;  CURRENT FILING DATE:  2002-03-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-006-130A-243

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::| :   |  : |:|              | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |    |    |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|          : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |   ::|                       |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||    | :    | | |  :     :  :   | : ::| ::  | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                              ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 5
US-10-199-672-310
```

```
; Sequence 310, Application US/10199672
; Publication No. US20030148442A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C1
;  CURRENT APPLICATION NUMBER: US/10/199,672
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: US/10/052,586
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059266
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/063120
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063121
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063486
;  PRIOR FILING DATE: 1997-10-21
;  PRIOR APPLICATION NUMBER: 60/063540
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063541
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063544
;  PRIOR FILING DATE: 1997-10-28
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-199-672-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:   :: | :| | |  ::  :     | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |      |     |       :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                  :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                  : :  | : | |      :||   :: ::| | :  : |       | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
               |         : : ||        | |    ::|                    |

Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
                |    ||||      | :      | | |  :      :  :    | : ::| :: | | |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
                :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                  ||  :    || :||    |  :   |: |  |
Db        402 --------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 6
US-10-006-172A-243
; Sequence 243, Application US/10006172A
; Publication No. US20030153000A1
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan l.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2830P1C11
;  CURRENT APPLICATION NUMBER: US/10/006,172A
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: 60/098716
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098723
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098749
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098750
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098803
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098821
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098843
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/099536
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099596
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099598
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099602
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099642
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099741
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099754
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099763
;  PRIOR FILING DATE: 1998-09-10
```

; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101071
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101279
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/101471
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101472
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101474
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102571
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102684
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102687
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102965
; PRIOR FILING DATE: 1998-10-02
; PRIOR APPLICATION NUMBER: 60/103258
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103314
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103315
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/105881
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/105882
; PRIOR FILING DATE: 1998-10-27

```
;  PRIOR APPLICATION NUMBER: 60/106023
;  PRIOR FILING DATE: 1998-10-28

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
               :  :|     :: |     |     |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |        | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |          : : ||       | |    ::|                        |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||      | :     | | |  :     :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
               :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG---------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                 || :   || :||    |  :   |: | |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 7
US-10-187-749-310
; Sequence 310, Application US/10187749
; Publication No. US20030153036A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C1
;  CURRENT APPLICATION NUMBER: US/10/187,749
;  CURRENT FILING DATE:  2002-07-01
;  PRIOR APPLICATION NUMBER: US/10/052,586
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059266
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/063120
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063121
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063486
;  PRIOR FILING DATE: 1997-10-21
;  PRIOR APPLICATION NUMBER: 60/063540
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063541
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063544
;  PRIOR FILING DATE: 1997-10-28
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-187-749-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
               :  :|     :: |     |     |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |        | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |          : : ||       | |    ::|                        |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||      | :     | | |  :     :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
               :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG---------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                 || :   || :||    |  :   |: | |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 8
US-10-194-457-310
; Sequence 310, Application US/10194457
; Publication No. US20030153037A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C296
;  CURRENT APPLICATION NUMBER: US/10/194,457
;  CURRENT FILING DATE:  2002-07-11
;  PRIOR APPLICATION NUMBER: 10/052586
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059266
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/063120
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063121
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063486
;  PRIOR FILING DATE: 1997-10-21
;  PRIOR APPLICATION NUMBER: 60/063540
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063541
;  PRIOR FILING DATE: 1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063544
;  PRIOR FILING DATE: 1997-10-28
;  Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-194-457-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |      ||         ::|  :   |  : |:|                 | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:   :: | :| | |  ::   :    | :  | ::      |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |     |    |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                  : :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |     :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
                |          : : ||        | |    ::|                    |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST----------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
                |   ||||    | :    | | |  :   :  :   | : ::| :: | | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
                :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG---------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                  ||  :   || :||   |  :   |: | |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 9
US-10-184-642-310
; Sequence 310, Application US/10184642
; Publication No. US20030157635A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C194
;  CURRENT APPLICATION NUMBER: US/10/184,642
;  CURRENT FILING DATE:  2002-06-27
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-642-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |      ||         ::|  :   |  : |:|                 | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |     |:   :: | :| | |  ::   :    | :  | ::      |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |     |    |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                  : :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |     :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
                |          : : ||        | |    ::|                    |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST----------------ATN 302
```

```
Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG------------------ 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 10
US-10-196-747-310
; Sequence 310, Application US/10196747
; Publication No. US20030162250A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C346
;  CURRENT APPLICATION NUMBER: US/10/196,747
;  CURRENT FILING DATE:  2002-07-16
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-196-747-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity  17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG-----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
```

```
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG------------------ 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 11
US-10-015-392A-243
; Sequence 243, Application US/10015392A
; Publication No. US20030166901A1
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan l.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2830P1C58
;  CURRENT APPLICATION NUMBER: US/10/015,392A
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: 60/098716
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098723
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098749
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098750
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098803
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098821
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098843
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/099536
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099596
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099598
;  PRIOR FILING DATE: 1998-09-09
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-015-392A-243

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity  17.7%;  Pred. No. 0.0046;
```

```
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |      ||        ::|  :    |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              :   :|     :: |     |    |     :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|           : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |    ::|                       |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||      | :     | | |  :     :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                            ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 12
US-10-017-253A-243
; Sequence 243, Application US/10017253A
; Publication No. US20030166055A1
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin P.
;  APPLICANT: Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan l.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2830P1C62
;  CURRENT APPLICATION NUMBER: US/10/017,253A
;  CURRENT FILING DATE:  2001-12-13
;  PRIOR APPLICATION NUMBER: 60/098716
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098723
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098749
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098750
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/098803
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098821
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/098843
;  PRIOR FILING DATE: 1998-09-02
;  PRIOR APPLICATION NUMBER: 60/099536
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099596
;  PRIOR FILING DATE: 1998-09-09
;  PRIOR APPLICATION NUMBER: 60/099598
;  PRIOR FILING DATE: 1998-09-09
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-017-253A-243

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |      ||        ::|  :    |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              :   :|     :: |     |    |     :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|           : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |    ::|                       |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||      | :     | | |  :     :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                            ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 13
US-10-173-689-310
; Sequence 310, Application US/10173689
; Publication No. US20030166104A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
```

```
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C10
;  CURRENT APPLICATION NUMBER: US/10/173,689
;  CURRENT FILING DATE:  2002-06-17
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-173-689-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |    ||         ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::      |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              :   :|     :: |      |     |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              : :  | ||  :    :   ::|||   | :::||::|            : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |   ::|                       |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||     | :      | | |  :     :  :   | : ::| :: | | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              ::  | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                 ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
              :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 14
US-10-173-690-310
; Sequence 310, Application US/10173690
; Publication No. US20030166105A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C9
;  CURRENT APPLICATION NUMBER: US/10/173,690
;  CURRENT FILING DATE:  2002-06-17
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-173-690-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |    ||         ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::      |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              :   :|     :: |      |     |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              : :  | ||  :    :   ::|||   | :::||::|            : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |   ::|                       |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
              |    ||||     | :      | | |  :     :  :   | : ::| :: | | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              ::  | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                 ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
              :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 15
US-10-173-691-310
; Sequence 310, Application US/10173691
; Publication No. US20030166106A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C13
;  CURRENT APPLICATION NUMBER: US/10/173,691
;  CURRENT FILING DATE:  2002-06-17
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-173-691-310

  Query Match             5.2%;  Score 148.5;  DB 12;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0046;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG-----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG------------------ 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


Search completed: December 16, 2003, 09:07:35
Job time : 29.2437 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 53.9828 Seconds
(without alignments)
1643.636 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPNVNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2851	100.0	559	20	AAY06221	EHEC E. coli trans
2	1498	52.5	549	20	AAY06220	EPEC E. coli trans
3	406	14.2	107	21	AAB20576	Intimin C-terminal
4	162.5	5.7	1300	22	ABB68075	Drosophila melanog
5	161	5.6	2586	22	ABB66878	Drosophila melanog
6	152.5	5.3	1412	22	ABB60943	Drosophila melanog
7	149.5	5.2	676	22	ABB62247	Drosophila melanog
8	148.5	5.2	596	21	AAY99408	Human PRO1342 (UNQ
9	148.5	5.2	596	22	AAU29178	Human PRO polypept
10	148.5	5.2	596	22	AAB87575	Human PRO1342. Ho
11	148.5	5.2	596	22	AAB66157	Protein of the inv
12	148.5	5.2	596	23	ABG95900	Human secreted/tra
13	148.5	5.2	596	24	ABU71266	Human PRO1342 prot
14	148.5	5.2	596	24	ABU71555	Human secreted pol
15	148.5	5.2	596	24	ABU72001	Novel human secret
16	148.5	5.2	596	24	ABU72158	Human PRO polypept
17	148.5	5.2	596	24	ABU65723	Human secreted/tra
18	148.5	5.2	596	24	ABU66056	Novel human secret
19	148.5	5.2	596	24	ABU67560	Human secreted/tra
20	148.5	5.2	596	24	ABU65418	Human PRO polypept
21	148.5	5.2	596	24	ABU58554	Human PRO polypept
22	148.5	5.2	596	24	ABU56090	Human secreted/tra
23	148.5	5.2	596	24	ABU57085	Human PRO polypept
24	148.5	5.2	596	24	ABU10664	Human secreted/tra
25	145	5.1	332	22	ABB67152	Drosophila melanog
26	145	5.1	1870	24	ABJ19019	Pathogen specific
27	142.5	5.0	914	22	ABB66394	Drosophila melanog
28	142.5	5.0	1026	22	ABB61775	Drosophila melanog
29	141.5	5.0	1086	22	ABB69356	Drosophila melanog
30	141.5	5.0	1959	12	AAR10562	Mutant protease (d
31	141.5	5.0	1962	12	AAR10560	Mutant protease (K
32	141.5	5.0	1962	12	AAR10561	Mutant protease (N
33	141.5	5.0	1962	12	AAR10557	Mutant protease (A
34	141.5	5.0	1962	12	AAR10558	Mutant protease (A
35	141.5	5.0	1962	12	AAR10559	Mutant protease (A
36	141.5	5.0	1962	12	AAR10563	Mutant protease (K
37	141.5	5.0	1968	12	AAR10941	Mutant protease (d
38	141.5	5.0	1974	12	AAR10940	Mutant protease (d
39	140.5	4.9	875	22	ABB71072	Drosophila melanog
40	140.5	4.9	1026	21	AAY83025	Staufen protein of
41	138	4.8	688	22	ABG50180	Human liver peptid
42	138	4.8	688	22	ABB30137	Peptide #2788 enco
43	138	4.8	688	22	ABB35307	Peptide #2813 enco
44	138	4.8	688	22	ABB20749	Protein #2748 enco
45	138	4.8	688	22	AAM56138	Human brain expres

ALIGNMENTS

RESULT 1
AAY06221
ID AAY06221 standard; Protein; 559 AA.
XX
AC AAY06221;
XX
DT 16-AUG-1999 (first entry)
XX
DE EHEC E. coli translocated intimin receptor (Tir).
XX
KW Tir; translocated intimin receptor; Hp90; enterohaemorrhagic;
KW EHEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Misc-difference 453
FT /note= "encoded by codon of 1 apparent nucleotide,
FT causing frameshift in the DNA sequence"
XX
PN WO9924576-A1.
XX
PD 20-MAY-1999.
XX
PF 10-NOV-1998; 98WO-CA01042.
XX
PR 12-NOV-1997; 97US-0065130.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Devinney R, Finlay BB, Kenny B, Stein M;
XX

DR WPI; 1999-337712/28.
DR N-PSDB; AAX58859.
XX
PT New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS Claim 7; Page 55-58; 91pp; English.
XX
CC The present sequence represents Tir, a novel translocated intimin
CC receptor (formerly termed Hp90) from an enterohaemorrhagic
CC Escherichia coli (EHEC) strain. The sequence was deduced from an
CC isolated tir polynucleotide (see AAX58859). Tir proteins are
CC secreted by attaching and effacing pathogens such as EHEC and EPEC
CC (see AAY06220) E. coli. The bacterial pathogens insert their own
CC receptors into mammalian cell surfaces, to which the pathogen then
CC adheres to trigger additional host signaling events and actin
CC nucleation. Diagnosis of disease caused by pathogenic E. coli can
CC be performed by use of antibodies that bind to Tir to detect the
CC protein or the use of nucleic acid probes for detection of nucleic
CC acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
CC peptides, a recombinant method for producing recombinant Tir,
CC antibodies which bind to Tir, and a kit for the detection of
CC Tir-producing E. coli are provided. A method of immunising a host
CC with Tir to induce a protective immune response is also provided.
CC In addition, Tir fusion proteins can be used in attenuated E. coli
CC to induce a cell-mediated immune response to other polypeptides,
CC e.g. antigens. A method for screening for compounds which
CC interfere with the binding of bacterial pathogens to their
CC receptors is further provided.
XX
SQ Sequence 559 AA;

Query Match 100.0%; Score 2851; DB 20; Length 559;
Best Local Similarity 100.0%; Pred. No. 4.4e-212;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGD 60

Qy         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA 120

Qy        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300

Qy        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSG 360

Qy        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQG 420

Qy        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 NVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRRCRLLILIRLF 480

Qy        481 RIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLTGGS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLALSGGLRHDMGGLTGGS 540

Qy        541 NSAVNTSNNPPAPGSHRFV 559
              ||||||||||||||||||||
Db        541 NSAVNTSNNPPAPGSHRFV 559
```

RESULT 2
AAY06220
ID AAY06220 standard; Protein; 549 AA.
XX
AC AAY06220;
XX
DT 16-AUG-1999 (first entry)
XX
DE EPEC E. coli translocated intimin receptor (Tir).
XX
KW Tir; translocated intimin receptor; Hp90; enteropathogenic;
KW EPEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Domain 234..253
FT /note= "putative transmembrane domain"
FT Domain 364..386
FT /note= "putative transmembrane domain"
FT Misc-difference 180
FT /note= "encoded by AAA"
FT Misc-difference 314
FT /note= "given as Xaa in the specification; Lys
FT is deduced from the DNA sequence"
XX
PN WO9924576-A1.
XX
PD 20-MAY-1999.
XX
PF 10-NOV-1998; 98WO-CA01042.
XX
PR 12-NOV-1997; 97US-0065130.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Devinney R, Finlay BB, Kenny B, Stein M;
XX
DR WPI; 1999-337712/28.
DR N-PSDB; AAX58858.
XX
PT New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS Claim 6; Page 55-58; 91pp; English.
XX
CC The present sequence represents Tir, a novel translocated intimin
CC receptor (formerly termed Hp90) from an enteropathogenic
CC Escherichia coli (EPEC) strain. The sequence was deduced from an
CC isolated tir polynucleotide (see AAX58858). Tir proteins are
CC secreted by attaching and effacing pathogens such as EPEC and EHEC
CC (see AAY06221) E. coli. The bacterial pathogens insert their own
CC receptors into mammalian cell surfaces, to which the pathogen then
CC adheres to trigger additional host signaling events and actin
CC nucleation. Diagnosis of disease caused by pathogenic E. coli can
CC be performed by use of antibodies that bind to Tir to detect the
CC protein or the use of nucleic acid probes for detection of nucleic
CC acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
CC peptides, a recombinant method for producing recombinant Tir,
CC antibodies which bind to Tir, and a kit for the detection of
CC Tir-producing E. coli are provided. A method of immunising a host
CC with Tir to induce a protective immune response is also provided.
CC In addition, Tir fusion proteins can be used in attenuated E. coli
CC to induce a cell-mediated immune response to other polypeptides,
CC e.g. antigens. A method for screening for compounds which
CC interfere with the binding of bacterial pathogens to their
CC receptors is further provided.
XX
SQ Sequence 549 AA;

Query Match 52.5%; Score 1498; DB 20; Length 549;
Best Local Similarity 54.8%; Pred. No. 2.4e-107;
Matches 323; Conservative 64; Mismatches 132; Indels 70; Gaps 12;

```
Qy          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58
              |||||||:| | |: |||||||||||||||  || | ||:||| ||||:||:|:|||||||
Db          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

Qy         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116
                |:|  |:|||| || ||||  ||  |  ||||||| |||| || ||| | |||| | ||
Db         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118

Qy        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176
              | |:|::||:| || || ||:: ||||| |||::|||||||||:|| |||||||| |||
Db        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178

Qy        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227
              |||  |:|   |           | | | : |:      : |    :|| |||: ||:||
Db        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230

Qy        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287
              ||| |:::| :| || ||||||| |||||||||| ||||||| ||:| |:||:||||:||
Db        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290

Qy        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347
              |:||:|||||||  |||||||  ||||:|  | :||| ||| |:|||:|:|||||::|:|
Db        291 FKNPENQKVNIDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350

Qy        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSART 407
              |:|||||::||| ||||| |||: ||||  || ||||:||| ||||||||          |
Db        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403

Qy        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS---------IGGPCRIR 458
              |  :    |      |         || |  | :|| :  ||          :||    |
Db        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGG---AR 460

Qy        459 MLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPP-----RDTTDNG---ARLLG 510
                :  |:                          :|  :   |     :: : :|    ||:|
Db        461 NSLSAHQ--------------------PEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIG 500

Qy        511 NPSAGIQSTYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559
              |  ||||||| || |||||  ||||| |  :||:: |  | ||  |||
Db        501 TPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549
```

RESULT 3
AAB20576
ID AAB20576 standard; protein; 107 AA.
XX
AC AAB20576;
XX
DT 08-DEC-2000 (first entry)
XX
DE Intimin C-terminal Tir binding domain amino acid sequence.
XX
KW Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
KW Tir-independent eukaryotic cell binding activity; bacterial infection;
KW diarrhoea; antibacterial.
XX
OS Unidentified.
XX
PN WO200045173-A1.
XX
PD 03-AUG-2000.
XX
PF 31-JAN-2000; 2000WO-GB00254.
XX
PR 29-JAN-1999; 99GB-0001897.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Frankel GM, Matthews SJ, Hale CB, Dougan G;
XX
DR WPI; 2000-499357/44.
XX
PT Screening for inhibitors of intimin binding to eukaryotic cells, for
PT use in diagnosing, preventing and treating bacterial infections,
PT especially Escherichia coli O157:H7 -
XX
PS Claim 8; Page 76; 96pp; English.
XX
CC The present invention describes a method of screening for an inhibitor
CC of intimin binding to eukaryotic cells. The method comprises exposing an
CC intimin polypeptide having a Tir-independent cell binding activity to
CC test agents, and obtaining an inhibitor based on its ability to bind the
CC polypeptide. The inhibitors are used in the prevention, treatment and/or
CC diagnosis of bacterial infections, preferably by enteropathic and/or
CC enterohaemorrhagic Escherichia coli, Shiga toxigenic E. coli, Hafnia
CC alvei or Citrobacter freundii, or especially E. coli O157:H7. The
CC infections cause a histopathological effect known as attachment and
CC effacement on intestinal epithelial cells. The inhibitors can be used
CC to produce food supplements or additives, especially where the food is
CC a milk substitute. The method can be used to sort cells based on their
CC ability to bind to a Tir independent cell binding domain of an intimin
CC polypeptide. Polypeptides having Tir-independent intimin binding
CC activity can be used to produce a vaccine against a bacterial disease.
CC The present sequence represents a specifically claimed intimin C-terminal
CC Tir binding domain amino acid sequence, for use in the method of the
CC present invention.
XX
SQ Sequence 107 AA;

Query Match 14.2%; Score 406; DB 21; Length 107;
Best Local Similarity 72.0%; Pred. No. 8.7e-24;
Matches 77; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

```
Qy        252 QALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLK 311
              |||||||||| ||||||| ||:| |:||:||||:|||:||:|||||||  ||||||| ||
Db          1 QALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIPSGELK 60

Qy        312 DDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVS 358
              ||:|  | :||| ||| |:|||:|:|||||::|::| |:|||||::|
Db         61 DDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLS 107
```

RESULT 4
ABB68075
ID ABB68075 standard; Protein; 1300 AA.
XX
AC ABB68075;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31017.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12178.

```
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 31017; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1300 AA;

  Query Match             5.7%;  Score 162.5;  DB 22;  Length 1300;
  Best Local Similarity   20.6%;  Pred. No. 0.0019;
  Matches   99;  Conservative   61;  Mismatches  189;  Indels  131;  Gaps   18;

Qy          5 NLGHNPNVNNSIPPAPPLPSQTDG---AGGRGQLINSTGPLGSRALFTPVRNSMADSGDN 61
              |   | | | :    |   | |:|   :|| |       |  |:  | ||   |   :|
Db        438 NSTSNSNSNTNDSTGPSETSSTNGLVASGGAG------GATGAAMLPTP---SQQSTGGK 488

Qy         62 RAS--------DVPGLPVNPMRLA-----------ASEITLNDGFEVLHDHGPLDTLNRQ 102
               |:         :| : |:|: :           |  |  |  ::  | |
Db        489 EATAAVSLLEKKLPNVVVSPLTMKELRQKGMTKYDAEMIMANAAYQQQHHH--------- 539

Qy        103 IGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAG 162
                      |          |   ||          : |: | :|             || | ||
Db        540 --QHHFHHHHHHHHHHHNHGQHASTGAEATAAVQQMAAMQKPG---------VGGTGAAG 588

Qy        163 HAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTST 222
              :|  |  | :                | |     | |  : :   ||          ||:|
Db        589 NAGATTVSSVA-----------AGAGSEVNGGRSTSLRKSMRVNS---------TSSSI 627

Qy        223 SSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQ 282
              |:  :|  :                  | | :::    :|    |    :    |   |
Db        628 STASADEVI----------------APVVAASISLPSKAPVVLMPRCKPAQMAIAALHQ 670

Qy        283 LTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQ---------A 333
                :   :  : ||  : :          |    |    : |:| |    | |||         :
Db        671 SQQRQLRRSERQKEKLTD-------GESSD---TSSEQQKK----EQKQQDHQLPQKMFS 716

Qy        334 IENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQT 393
              :    | :|  :|:| ::|: :  :|    ||          ||  |  |  |
Db        717 LAEEPQPEKS-EEKQQEQQKRVTRNSAGRVGL-----------VARLATAHNNN----IA 760

Qy        394 TTTTTTTTTTSARTV---ENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS 450
              ||| :::::  | |:    |  :||:    :     |  :::||: : :  ||    |||
Db        761 TTTNSSSSSNKATTITNCNNHNSNNSSRINHNSNLSSRLSVKSRKPAPSEASSIPSSTSS 820


RESULT 5
ABB66878
ID   ABB66878 standard; Protein; 2586 AA.
XX
AC   ABB66878;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 27426.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10981.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 27426; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2586 AA;

  Query Match             5.6%;  Score 161;  DB 22;  Length 2586;
  Best Local Similarity   19.8%;  Pred. No. 0.0066;
  Matches   99;  Conservative   69;  Mismatches  209;  Indels  122;  Gaps   17;

Qy          8 HNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRASDVP 67
              | | | |  |   |:  : | |       |:         :    : :| |   :|
Db       1123 HLPMVANPPPHLTPVVEVSQGTNGDNSSTQSSS--------STTTTSSDEGQTTSSSA- 1173

Qy         68 GLPVNPMRLAASEITLNDGFEVLHDHGPL---------DTLNRQIGSSVFRVETQEDGKH 118
                ||:      : | :|| :    |:         |  : |  ||     : ::|:
Db       1174 --PVSSTTTTTTTTTSSDGDQSTTSSDPVVEVSQGTNGDNSSTQSSSSTTTTSSDEGQT 1231

Qy        119 IAV-------------GQRNGVE------TSVVLSDQEYARLQSIDPEGKDKFVFTGGR 158
               :                |  |  :      |:  |         | ||  :   |  |
Db       1232 TSSSAPVVEVTQGSSSNGDGNSTQSSTTTTTTTTSSDGGESTTSSDPVVE---VSQGTN 1288

Qy        159 G------GAGHAMVTVASD---ITEARQRILELLEPKGTGESKGAGESKGVGELRESNSG 209
              |      :     | :||    | :   ::|: :             | | |   :|::
Db       1289 GDNSSTQSSSSTTTTSSDEGQTTSSSAPVVEVTQ---------GSSSNGDGNSTQSSTT 1339

Qy        210 AENTTETQTSTSTSSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPD 269
                || | :    |:  ||| : :: ||                          |: :|
Db       1340 TTTTTTTSSDGGESTTSSDPVVEVSQGT--------------------NGDNSST---Q 1375

Qy        270 AAASATETATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVL------KDDVV---ANIEE 320
              :::| | | : |:    :   |      : ::    | ||        ||:   ||
Db       1376 SSSSTTTTSSDEGQTTSSSGP------VVDISQGSSSNVLDNDNNYNDDIFRWWANPPP 1429

Qy        321 QAKAAGEEAKQQAIENNAQAQKKYDEQQAKRQEELKVSSGA-----GYGLSGALILGGGI 375
                  :  |:|       : ::  :||     |    || |        | |       |
Db       1430 YLTPLWKSVKEQMAITAPLSLRRPPQQQTSSDEGQTTSSSAPVVDISQGSS-----SNGD 1484

Qy        376 GVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVENKPANNTPAQGNVDTPGSEDTMESRR 435
              | :             |::||||||||||:   |:   ::   :  | |  :: :|
Db       1485 GNST-----------QSSTTTTTTTTSSDGGESTTLSDPVVEVSQGTNGDNNSTQSSS 1532

Qy        436 SSMASTSSTFFDTSSIGGP 454
```

```
Db       1533 STTTTTSSDEGQTTSSSDP 1551
```

RESULT 6
ABB60943
ID ABB60943 standard; Protein; 1412 AA.
XX
AC ABB60943;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 9621.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05046.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 9621; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1412 AA;

Query Match 5.3%; Score 152.5; DB 22; Length 1412;
Best Local Similarity 19.4%; Pred. No. 0.013;
Matches 139; Conservative 100; Mismatches 256; Indels 223; Gaps 31;

```
Qy          13 NNSIPP--------APPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADS-GDNRA 63
Db         684 SNSASPAKFEVPVRASPSPAKSD--------------TSSHASTSRTRNSPAPSPGKFSV 729

Qy          64 SDV----PGLPVNPMRLAASEI---------TLND----GFEVLHDHGPLDTLNRQIGSS 106
Db         730 SDIARNSPSITPNKYEAAAAALVPPAAACLPTANDLMAASLAQLAGQLPPNFLQGDL-AA 788

Qy         107 VFRVETQEDGK-----------HIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFT 155
Db         789 LFQQQRKDYGSSSLNQLAAAAAKVGGSKQSNASASSGLNDPNVAAAVAAYSNSFNMPLPT 848

Qy         156 G-GRGGAGHAMVTVASDITEARQRILELLEPKGTGESKGAGESKGVGELRESNSGA-ENT 213
Db         849 GVGIGGSGNS------------------NAHSSSKSKSERSSKSSSSSSSSNSSSTSNS 889

Qy         214 TETQTSTSTSSLRSDP-------------KLWL----ALGTVATGLIGL---------AA 247
Db         890 YKTKLMKELDELKNDPLKMSELIRSPEYAALLLQQAEALGATTLGTLGFGSDYSYLTGAG 949

Qy         248 TGIVQALALTPEPDSPTT------TDPDAAASATETATRDQLTK-------EAFQNPDNQ 294
Db         950 LGVPAANALTGGQSSNSSSSKSSKSSPAAAAAAALSADYNNLIQASKLLGYDSYMQQSKQ 1009

Qy         295 KVNID---ELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
Db        1010 SNDLNAFLQQQMAVAAASIPPPPQTASSGSSNSSSSKKQQQLQQQQHQQQQQQQQQAAA 1069

Qy         352 QEEL-----------------------KVSSGAGYGLSGALILGGGIG---------- 376
Db        1070 QADYTALLQTYTKLFDPNNQFAAAMSSNKHMAGAHNELSALLSSGVGVGGGASGGGSKQK 1129

Qy         377 -------------------VAVTAALHRKNQPVEQ------TTTTTTTTTTSARTVENK 411
Db        1130 QKDIQSDMLNQLLQLEKQDSEIKALLYRQNKAAADLDALFATPSGAVVGAGSSANAMKSG 1189

Qy         412 PANNTPAQGNVDTPGSEDTMESRRSSMASTS----STFFDTS----SIGGPCRIRMLMLK 463
Db        1190 SSAGNSGVSGMSSPSSLSNQAAYYNALAQEKMQDYAAFFQQQHGKYGIPDPLSKTTLAAN 1249

Qy         464 HRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGI--QSTYA 521
Db        1250 NMFMTPSA-------LFKIQQESLSAM---MMKPPKSTTPSSARTRESSASPALERLTPT 1299

Qy         522 RLALSGGLRHDMGGLTGGSNS-------------AVNT--SNNP-PAP------GSHR 557
Db        1300 KSASSGG----SGG--GGSNSNSGGKYNFSAVDLAISSVPSNTPSPAPSDGSSGSSHR 1351
```

RESULT 7
ABB62247
ID ABB62247 standard; Protein; 676 AA.
XX
AC ABB62247;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13533.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06350.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 13533; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent

```
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   676 AA;

  Query Match             5.2%;  Score 149.5;  DB 22;  Length 676;
  Best Local Similarity   19.3%;  Pred. No. 0.008;
  Matches  103;  Conservative   68;  Mismatches  187;  Indels  177;  Gaps   17;

Qy         63 ASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIA-- 120
              || | | |    ::||    | :|: :      ||||   :| |:        | |:|
Db        150 ASSVAGAPGAGKKVAALVDFLVEGYGM-----SLDTL-EIVGFSL--------GAHVAGH 195

Qy        121 ---------VGQRNGVE----------TSVVLSDQEYARLQSI----------DPEGKDK 151
                      ||:  |::          |   ||  :   ::||           | ||
Db        196 TAKQVNSGKVGKVVGLDPASPLISYSNTEKRLSSDDALYVESIQTNGAILGFGQPIGKAS 255

Qy        152 FVFTGGRGGAGHAM-VTVASDITEARQRILELL-------------------------- 183
              |   |||    |  : :|  :   |:|    :| |
Db        256 FYMNGGRSQPGCGIDITGSCSHTKAVLYYVEALLWNNFPSIKCESSVDANKNNCGNTYSS 315

Qy        184 ----------------------EPKGTGESKGAGE-SKGVGELRESNSGAENTTETQTS 219
                                   | | ||    || : |  |:  :    |: : |: |
Db        316 VFMGASINFFVAEGIFYVPVNKESPYGLGELNSGGEATTGTPEITSTTVDGEDESTTEVS 375

Qy        220 TSTSSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATE--- 276
              |||:: : :                    ::    || || |   |: ::: ||
Db        376 TSTTTDKPE----------------------ESTTEEPEEDSTTNGKPEESSTTTEQPE 412

Qy        277 --TATRDQLTKEAFQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAA---GEEAKQ 331
                | | ::      : |:::           |   : :      |:   :    |::
Db        413 DSTTTTEEPIDSTTEAPEDESTTSSPTDGGEQSTTEEPETTTEKPEETSTSPIDTEDSTT 472

Qy        332 QAIENNAQAQKKYDEQQAKRQEELKVSSGAGYGLSGALILGGGIGVAVT----------- 380
              :  |: :   |: :|      |:   :|  |          ||      |
Db        473 KEPEDVSTTPKESEESTTSSPEDDDTTSAPGEDDDTTAAPGGDEEETTTEDPEETTTSPS 532

Qy        381 -AALHRKNQPVEQTTTT-------TTT----------------------TTTTSARTVEN 410
               |      :| | ||||      |||                      |:||:|
Db        533 DADDSTTEEPEEDTTTTKKPVEPSTTTEEPEDSTTEVPEESTTEEPDKETSTTNA----- 587

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDT---SSIGGPCRIRMLML 462
              :|   |  : :||:    :  :| |  :   |:|   |   |: | |  : :  |
Db        588 EPEVTTTVEPDVDSTVDPEGDQSTREDLEDVSTTAVPTKLPSTTGVPPEVPITTL 642


RESULT 8
AAY99408
ID   AAY99408 standard; Protein; 596 AA.
XX
AC   AAY99408;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1342 (UNQ697) amino acid sequence SEQ ID NO:243.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
PR   10-SEP-1998;   98US-0099763.
PR   10-SEP-1998;   98US-0099792.
PR   10-SEP-1998;   98US-0099808.
PR   10-SEP-1998;   98US-0099812.
PR   10-SEP-1998;   98US-0099815.
PR   10-SEP-1998;   98US-0099816.
PR   15-SEP-1998;   98US-0100385.
PR   15-SEP-1998;   98US-0100388.
PR   15-SEP-1998;   98US-0100390.
PR   16-SEP-1998;   98US-0100584.
PR   16-SEP-1998;   98US-0100627.
PR   16-SEP-1998;   98US-0100661.
PR   16-SEP-1998;   98US-0100662.
PR   16-SEP-1998;   98US-0100664.
PR   17-SEP-1998;   98US-0100683.
PR   17-SEP-1998;   98US-0100684.
PR   17-SEP-1998;   98US-0100710.
PR   17-SEP-1998;   98US-0100711.
PR   17-SEP-1998;   98US-0100919.
PR   17-SEP-1998;   98US-0100930.
PR   18-SEP-1998;   98US-0100848.
PR   18-SEP-1998;   98US-0100849.
PR   18-SEP-1998;   98US-0101014.
PR   18-SEP-1998;   98US-0101068.
PR   18-SEP-1998;   98US-0101071.
PR   22-SEP-1998;   98US-0101279.
PR   23-SEP-1998;   98US-0101471.
PR   23-SEP-1998;   98US-0101472.
PR   23-SEP-1998;   98US-0101474.
PR   23-SEP-1998;   98US-0101475.
PR   23-SEP-1998;   98US-0101476.
PR   23-SEP-1998;   98US-0101477.
PR   23-SEP-1998;   98US-0101479.
PR   24-SEP-1998;   98US-0101738.
PR   24-SEP-1998;   98US-0101741.
PR   24-SEP-1998;   98US-0101743.
PR   24-SEP-1998;   98US-0101915.
PR   24-SEP-1998;   98US-0101916.
PR   29-SEP-1998;   98US-0102207.
PR   29-SEP-1998;   98US-0102240.
PR   29-SEP-1998;   98US-0102307.
PR   29-SEP-1998;   98US-0102330.
PR   29-SEP-1998;   98US-0102331.
PR   30-SEP-1998;   98US-0102484.
PR   30-SEP-1998;   98US-0102487.
PR   30-SEP-1998;   98US-0102570.
PR   30-SEP-1998;   98US-0102571.
PR   01-OCT-1998;   98US-0102684.
PR   01-OCT-1998;   98US-0102687.
PR   02-OCT-1998;   98US-0102965.
PR   06-OCT-1998;   98US-0103258.
PR   06-OCT-1998;   98US-0103449.
PR   07-OCT-1998;   98US-0103314.
PR   07-OCT-1998;   98US-0103315.
PR   07-OCT-1998;   98US-0103328.
PR   07-OCT-1998;   98US-0103395.
PR   07-OCT-1998;   98US-0103396.
PR   07-OCT-1998;   98US-0103401.
```

```
PR   08-OCT-1998;   98US-0103633.
PR   08-OCT-1998;   98US-0103678.
PR   08-OCT-1998;   98US-0103679.
PR   08-OCT-1998;   98US-0103711.
PR   14-OCT-1998;   98US-0104257.
PR   20-OCT-1998;   98US-0104987.
PR   20-OCT-1998;   98US-0105000.
PR   20-OCT-1998;   98US-0105002.
PR   21-OCT-1998;   98US-0105104.
PR   22-OCT-1998;   98US-0105169.
PR   22-OCT-1998;   98US-0105266.
PR   26-OCT-1998;   98US-0105693.
PR   26-OCT-1998;   98US-0105694.
PR   27-OCT-1998;   98US-0105807.
PR   27-OCT-1998;   98US-0105881.
PR   27-OCT-1998;   98US-0105882.
PR   27-OCT-1998;   98US-0106062.
PR   28-OCT-1998;   98US-0106023.
PR   28-OCT-1998;   98US-0106029.
PR   28-OCT-1998;   98US-0106030.
PR   28-OCT-1998;   98US-0106032.
PR   28-OCT-1998;   98US-0106033.
PR   28-OCT-1998;   98US-0106178.
PR   29-OCT-1998;   98US-0106248.
PR   29-OCT-1998;   98US-0106384.
PR   29-OCT-1998;   98US-0108500.
PR   30-OCT-1998;   98US-0106464.
PR   03-NOV-1998;   98US-0106856.
PR   03-NOV-1998;   98US-0106902.
PR   03-NOV-1998;   98US-0106905.
PR   03-NOV-1998;   98US-0106919.
PR   03-NOV-1998;   98US-0106932.
PR   03-NOV-1998;   98US-0106934.
PR   10-NOV-1998;   98US-0107783.
PR   17-NOV-1998;   98US-0108775.
PR   17-NOV-1998;   98US-0108779.
PR   17-NOV-1998;   98US-0108787.
PR   17-NOV-1998;   98US-0108788.
PR   17-NOV-1998;   98US-0108801.
PR   17-NOV-1998;   98US-0108802.
PR   17-NOV-1998;   98US-0108806.
PR   17-NOV-1998;   98US-0108807.
PR   17-NOV-1998;   98US-0108867.
PR   17-NOV-1998;   98US-0108925.
PR   18-NOV-1998;   98US-0108848.
PR   18-NOV-1998;   98US-0108849.
PR   18-NOV-1998;   98US-0108850.
PR   18-NOV-1998;   98US-0108851.
PR   18-NOV-1998;   98US-0108852.
PR   18-NOV-1998;   98US-0108858.
PR   18-NOV-1998;   98US-0108904.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-237871/20.
DR   N-PSDB; AAA37090.
XX
PT   New mammalian DNA sequences encoding transmembrane, receptor or
PT   secreted PRO polypeptides, useful for screening of potential peptide or
PT   small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS   Claim 12; Fig 138; 773pp; English.
XX
CC   AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC   receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC   transmembrane and receptor PRO proteins can be used for screening of
CC   potential peptide or small molecule inhibitors of the relevant
CC   receptor/ligand interactions. The polypeptides and nucleotide sequences
CC   encoding then have various industrial applications, including uses as
CC   pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC   PCR primers and hybridisation probes used in the isolation of the PRO
CC   polypeptides from the present invention.
XX
SQ   Sequence   596 AA;

  Query Match             5.2%;  Score 148.5;  DB 21;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0081;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |      ||         ::| :   |  : |:|                 | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG---------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |      |    |       :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                  : :  | ||  :    :   ::|||   | :::||::|               : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG-----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                  : :  | : | |       :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
               |         : : ||         | |    ::|                      |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
               |    ||||     | :    | | |  :    :  :   | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
               :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                  ||  :    || :||    |  :   |: |  |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 9
AAU29178
ID   AAU29178 standard; Protein; 596 AA.
XX
AC   AAU29178;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human PRO polypeptide sequence #155.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000; 2000WO-US05601.
```

```
PR   02-MAR-2000; 2000WO-US05841.
PR   03-MAR-2000; 2000US-187202P.
PR   06-MAR-2000; 2000US-186968P.
PR   14-MAR-2000; 2000US-189320P.
PR   14-MAR-2000; 2000US-189328P.
PR   15-MAR-2000; 2000WO-US06884.
PR   21-MAR-2000; 2000US-190828P.
PR   21-MAR-2000; 2000US-191007P.
PR   21-MAR-2000; 2000US-191048P.
PR   21-MAR-2000; 2000US-191314P.
PR   28-MAR-2000; 2000US-192655P.
PR   29-MAR-2000; 2000US-193032P.
PR   29-MAR-2000; 2000US-193053P.
PR   30-MAR-2000; 2000WO-US08439.
PR   04-APR-2000; 2000US-194449P.
PR   04-APR-2000; 2000US-194647P.
PR   11-APR-2000; 2000US-195975P.
PR   11-APR-2000; 2000US-196000P.
PR   11-APR-2000; 2000US-196187P.
PR   11-APR-2000; 2000US-196690P.
PR   11-APR-2000; 2000US-196820P.
PR   18-APR-2000; 2000US-198121P.
PR   18-APR-2000; 2000US-198585P.
PR   25-APR-2000; 2000US-199397P.
PR   25-APR-2000; 2000US-199550P.
PR   25-APR-2000; 2000US-199654P.
PR   03-MAY-2000; 2000US-201516P.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   05-JUN-2000; 2000US-209832P.
PR   28-JUL-2000; 2000WO-US20710.
PR   22-AUG-2000; 2000US-0644848.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-602746/68.
DR   N-PSDB; AAS46079.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PS   Claim 11; Fig 310; 774pp; English.
XX
CC   Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC   The PRO polypeptides and their associated nucleic acids can be used to
CC   detect the presence of a tumour in a mammal by comparing the level of
CC   expression of a PRO polypeptide in a test sample of cells from the animal
CC   and a control sample of normal cells, whereby a higher level of
CC   expression in the test sample indicates the presence of a tumour in the
CC   mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC   and rabbits but are preferably human. The polypeptides can be used to
CC   stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC   when contacted with it. A specific polypeptide can be used to stimulate
CC   the proliferation or differentiation of chondrocyte cells. The PRO
CC   proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence   596 AA;

  Query Match             5.2%;  Score 148.5;  DB 22;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0081;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | : |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
                | |    |:   :: | :| | |  ::  :    | :  | ::      |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
               :   :|     :: |      |     |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG-----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
               |         : : ||      | |    ::|                      |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
               |    ||||      | :     | | |  :     :  :   | : ::| :: | | |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
                 :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                  ||  :    || :||    |  :   |: | |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 10
AAB87575
ID   AAB87575 standard; Protein; 596 AA.
XX
AC   AAB87575;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO1342.
XX
KW   Human; PRO protein; mapping.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
```

```
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92107.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 100; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the PRO protein,
CC   agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC   employed as molecular weight markers for protein electrophoresis. The PRO
CC   coding sequence has applications in molecular biology, including use as
CC   hybridisation probes, and in chromosome and gene mapping.
XX
SQ   Sequence   596 AA;

  Query Match             5.2%;  Score 148.5;  DB 22;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0081;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |  |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
               :   :|     :: |      |     |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|              : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |          | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
              |         : : ||        | |    ::|                        |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
              |    ||||     | :     | | |  :     :  :    | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
               :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                 ||  :    || :||    |  :   |: | |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:       :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477
```

```
RESULT 11
AAB66157
ID   AAB66157 standard; protein; 596 AA.
XX
AC   AAB66157;
XX
DT   02-APR-2001  (first entry)
XX
DE   Protein of the invention #69.
XX
KW   Secreted; transmembrane; gene therapy.
XX
OS   Unidentified.
XX
PN   WO200078961-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   16-DEC-1999;   99WO-US30095.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N,  Fong S;
PI   Gao W,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Pan J,  Paoni NF,  Roy MA,  Smith V,  Stewart TA,  Tumas D;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 1; Fig 138; 787pp; English.
XX
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti-sense RNA and DNA. They may also be used used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence   596 AA;

  Query Match             5.2%;  Score 148.5;  DB 22;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0081;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |  |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
               :   :|     :: |      |     |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163
```

```
Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG------------------ 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477
```

RESULT 12
ABG95900
ID ABG95900 standard; Protein; 596 AA.
XX
AC ABG95900;
XX
DT 10-DEC-2002 (first entry)
XX
DE Human secreted/transmembrane protein PRO1342.
XX
KW Human; secreted protein; transmembrane protein; antirheumatic;
KW antiarthritic; osteopathic; sports-related joint problem;
KW articular cartilage defect; osteoarthritis; rheumatoid arthritis.
XX
OS Homo sapiens.
XX
PN US2002119130-A1.
XX
PD 29-AUG-2002.
XX
PF 06-DEC-2001; 2001US-0006867.
XX
PR 29-OCT-1997; 97US-063435P.
PR 29-OCT-1997; 97US-064215P.
PR 22-APR-1998; 98US-082797P.
PR 29-APR-1998; 98US-083495P.
PR 15-MAY-1998; 98US-085579P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 11-JUN-1998; 98US-088863P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089514P.
PR 16-SEP-1998; 98WO-US19330.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21194.
PR 22-DEC-1999; 99WO-US30720.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 30-MAR-2000; 2000WO-US08439.
PR 22-MAY-2000; 2000WO-US14042.
PR 02-JUN-2000; 2000WO-US15264.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32378.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
XX
PA (GETH) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2002-731348/79.
DR N-PSDB; ABS74427.
XX
PT New isolated secreted and transmembrane PRO polypeptide useful for
PT modulating biological activity of a cell, or for treating
PT sports-related joint problems, osteoarthritis or rheumatoid arthritis
XX
PS Claim 20; Fig 100; 399pp; English.
XX
CC The invention relates to an isolated secreted and transmembrane PRO
CC polypeptide having 80 % sequence identity to a sequence appearing
CC as ABG95851-ABG95934 or their associated signal peptide, or a sequence of
CC an extracellular domain of the proteins with their associated signal
CC peptide or lacking its associated signal peptide. Also included are
CC the nucleic acids encoding the proteins, vectors, host cells,
CC fusion proteins and antibodies which specifically bind to the proteins.
CC The proteins are useful for detecting a polypeptide designated as A, B, C
CC or D in a sample suspected of containing an A, B, C or D polypeptide,
CC by contacting the sample with a polypeptide designated as E, F, G, H or
CC I (or vice versa) and determining the formation of a A/E, B/F, B/G, C/H
CC or D/I polypeptide conjugate in the sample, where the formation of the
CC conjugate is indicative of the presence of an A, B, C or D polypeptide
CC in the sample, where A is a PRO10272 polypeptide, B is a PRO20110
CC polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,
CC E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040
CC polypeptide, H is a PRO20233 polypeptide and I is a PRO1890
CC polypeptide. The sample comprises a cell suspected of expressing the A,
CC B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with
CC a detectable label or is attached to a solid support. The proteins are
CC useful for linking a bioactive molecule to a cell expressing a
CC polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive
CC molecule is a toxin, a radiolabel or an antibody. The bioactive molecule
CC causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies
CC against them are useful for modulating a biological activity of a cell
CC expressing a polypeptide designated as A, B, C or D or E, F, G, H, or
CC I. The cell is killed. The proteins are useful for identifying
CC agonists or antagonists, for the preparation of a medicament useful in
CC the treatment of a condition which is responsive to the proteins, as
CC molecular weight markers for protein electrophoresis purposes, and as
CC therapeutic agents for treating sports-related joint problems,
CC articular cartilage defects, osteoarthritis or rheumatoid arthritis.
CC Nucleic acids encoding the proteins are useful as hybridisation probes,
CC in chromosome and gene mapping, in the generation of anti-sense RNA and
CC DNA, for the preparation of the proteins, to generate transgenic or
CC knockout animals which are useful in the development and screening of
CC therapeutic useful reagents, for chromosome identification, and in gene
CC therapy. The antibody is useful as a therapeutic agent, in a diagnostic
CC assay and for affinity purification of the protein from recombinant
CC cell culture natural sources. The present sequence represents a novel
CC secreted or transmembrane protein of the invention.
XX
SQ Sequence 596 AA;

Query Match 5.2%; Score 148.5; DB 23; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0081;

```
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
Db        402 --------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 13
ABU71266
ID   ABU71266 standard; Protein; 596 AA.
XX
AC   ABU71266;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human PRO1342 protein.
XX
KW   Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood;
KW   tumour necrosis factor alpha release; chondrocyte cell; proliferation;
KW   differentiation; tumour; gene therapy.
XX
OS   Homo sapiens.
XX
PN   US2003036143-A1.
XX
PD   20-FEB-2003.
XX
PF   02-JUL-2002; 2002US-0187600.
XX
PR   16-SEP-1998;   98WO-US19330.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   02-JUN-1999;   99WO-US12252.
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   01-DEC-1999;   99WO-US28301.
PR   02-DEC-1999;   99WO-US28551.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   29-AUG-2001; 2001WO-US27099.
PR   18-SEP-1997;   97US-059263P.
PR   18-SEP-1997;   97US-059266P.
PR   17-OCT-1997;   97US-062250P.
PR   21-OCT-1997;   97US-063486P.
PR   24-OCT-1997;   97US-063120P.
PR   24-OCT-1997;   97US-063121P.
PR   28-OCT-1997;   97US-063540P.
PR   28-OCT-1997;   97US-063541P.
PR   28-OCT-1997;   97US-063544P.
PR   28-OCT-1997;   97US-063564P.
PR   29-OCT-1997;   97US-063734P.
PR   31-OCT-1997;   97US-063870P.
PR   31-OCT-1997;   97US-064103P.
PR   13-NOV-1997;   97US-065311P.
PR   21-NOV-1997;   97US-066120P.
PR   24-NOV-1997;   97US-066466P.
PR   24-NOV-1997;   97US-066772P.
PR   11-DEC-1997;   97US-069335P.
PR   12-DEC-1997;   97US-069425P.
PR   17-DEC-1997;   97US-069870P.
PR   18-DEC-1997;   97US-068017P.
PR   10-MAR-1998;   98US-077450P.
PR   11-MAR-1998;   98US-077632P.
PR   11-MAR-1998;   98US-077649P.
PR   20-MAR-1998;   98US-078886P.
PR   20-MAR-1998;   98US-078939P.
PR   27-MAR-1998;   98US-079664P.
PR   27-MAR-1998;   98US-079786P.
PR   31-MAR-1998;   98US-080107P.
PR   31-MAR-1998;   98US-080194P.
PR   01-APR-1998;   98US-080327P.
PR   01-APR-1998;   98US-080333P.
PR   08-APR-1998;   98US-081049P.
PR   08-APR-1998;   98US-081070P.
PR   09-APR-1998;   98US-081195P.
PR   15-APR-1998;   98US-081838P.
PR   21-APR-1998;   98US-082568P.
PR   21-APR-1998;   98US-082569P.
PR   22-APR-1998;   98US-082704P.
PR   22-APR-1998;   98US-082797P.
PR   28-APR-1998;   98US-083322P.
PR   29-APR-1998;   98US-083495P.
PR   29-APR-1998;   98US-083496P.
PR   29-APR-1998;   98US-083499P.
PR   29-APR-1998;   98US-083559P.
PR   05-MAY-1998;   98US-084366P.
```

```
PR   06-MAY-1998;   98US-084414P.
PR   07-MAY-1998;   98US-084639P.
PR   07-MAY-1998;   98US-084640P.
PR   07-MAY-1998;   98US-084643P.
PR   15-MAY-1998;   98US-085579P.
PR   15-MAY-1998;   98US-085580P.
PR   15-MAY-1998;   98US-085582P.
PR   15-MAY-1998;   98US-085700P.
PR   18-MAY-1998;   98US-086023P.
PR   22-MAY-1998;   98US-086392P.
PR   22-MAY-1998;   98US-086486P.
PR   28-MAY-1998;   98US-087098P.
PR   28-MAY-1998;   98US-087208P.
PR   02-JUN-1998;   98US-087609P.
PR   02-JUN-1998;   98US-087759P.
PR   03-JUN-1998;   98US-087827P.
PR   04-JUN-1998;   98US-088025P.
PR   04-JUN-1998;   98US-088028P.
PR   04-JUN-1998;   98US-088029P.
PR   04-JUN-1998;   98US-088033P.
PR   04-JUN-1998;   98US-088326P.
PR   05-JUN-1998;   98US-088167P.
PR   05-JUN-1998;   98US-088202P.
PR   05-JUN-1998;   98US-088212P.
PR   05-JUN-1998;   98US-088217P.
PR   09-JUN-1998;   98US-088655P.
PR   10-JUN-1998;   98US-088722P.
PR   10-JUN-1998;   98US-088738P.
PR   10-JUN-1998;   98US-088740P.
PR   10-JUN-1998;   98US-088811P.
PR   10-JUN-1998;   98US-088824P.
PR   10-JUN-1998;   98US-088825P.
PR   10-JUN-1998;   98US-088826P.
PR   11-JUN-1998;   98US-088861P.
PR   11-JUN-1998;   98US-088863P.
PR   11-JUN-1998;   98US-088876P.
PR   12-JUN-1998;   98US-089090P.
PR   12-JUN-1998;   98US-089105P.
PR   16-JUN-1998;   98US-089512P.
PR   16-JUN-1998;   98US-089514P.
PR   17-JUN-1998;   98US-089538P.
PR   17-JUN-1998;   98US-089598P.
PR   17-JUN-1998;   98US-089653P.
PR   18-JUN-1998;   98US-089908P.
PR   19-JUN-1998;   98US-089952P.
PR   22-JUN-1998;   98US-090246P.
PR   22-JUN-1998;   98US-090252P.
PR   22-JUN-1998;   98US-090254P.
PR   24-JUN-1998;   98US-090429P.
PR   24-JUN-1998;   98US-090435P.
PR   24-JUN-1998;   98US-090444P.
PR   24-JUN-1998;   98US-090461P.
PR   24-JUN-1998;   98US-090535P.
PR   24-JUN-1998;   98US-090540P.
PR   25-JUN-1998;   98US-090676P.
PR   25-JUN-1998;   98US-090678P.
PR   25-JUN-1998;   98US-090688P.
PR   25-JUN-1998;   98US-090690P.
PR   25-JUN-1998;   98US-090694P.
PR   25-JUN-1998;   98US-090695P.
PR   25-JUN-1998;   98US-090696P.
PR   26-JUN-1998;   98US-090862P.
PR   26-JUN-1998;   98US-090863P.
PR   26-JUN-1998;   98US-091010P.
PR   01-JUL-1998;   98US-091359P.
PR   01-JUL-1998;   98US-091544P.
PR   02-JUL-1998;   98US-091478P.
PR   02-JUL-1998;   98US-091486P.
PR   02-JUL-1998;   98US-091626P.
PR   02-JUL-1998;   98US-091628P.
PR   02-JUL-1998;   98US-091632P.
PR   24-JUL-1998;   98US-094006P.
PR   04-AUG-1998;   98US-095282P.
PR   10-AUG-1998;   98US-095998P.
PR   10-AUG-1998;   98US-096012P.
PR   17-AUG-1998;   98US-096757P.
PR   17-AUG-1998;   98US-096766P.
PR   17-AUG-1998;   98US-096867P.
PR   17-AUG-1998;   98US-096891P.
PR   17-AUG-1998;   98US-096897P.
PR   18-AUG-1998;   98US-096949P.
PR   18-AUG-1998;   98US-096959P.
PR   18-AUG-1998;   98US-097022P.
PR   26-AUG-1998;   98US-097952P.
PR   26-AUG-1998;   98US-097954P.
PR   26-AUG-1998;   98US-097955P.
PR   26-AUG-1998;   98US-097971P.
PR   26-AUG-1998;   98US-097974P.
PR   26-AUG-1998;   98US-098014P.
PR   01-SEP-1998;   98US-098716P.
PR   01-SEP-1998;   98US-098723P.
PR   02-SEP-1998;   98US-098803P.
PR   02-SEP-1998;   98US-098821P.
PR   02-SEP-1998;   98US-098843P.
PR   09-SEP-1998;   98US-099602P.
PR   10-SEP-1998;   98US-099741P.
PR   10-SEP-1998;   98US-099754P.
PR   10-SEP-1998;   98US-099763P.
PR   10-SEP-1998;   98US-099812P.

  Query Match             5.2%;  Score 148.5;  DB 24;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0081;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG---------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST-------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG------------------ 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
Db        402 --------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477

RESULT 14
```

```
ABU71555
ID   ABU71555 standard; Protein; 596 AA.
XX
AC   ABU71555;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Human secreted polypeptide PRO1342.
XX
KW   Human; gene therapy; tumour; cancer.
XX
OS   Homo sapiens.
XX
PN   US2003013855-A1.
XX
PD   16-JAN-2003.
XX
PF   03-MAY-2002; 2002US-0063616.
XX
PR   30-DEC-1998;   98KR-0062142.
PR   08-MAR-1999;   99WO-US05028.
PR   14-MAY-1999;   99WO-US10733.
PR   30-DEC-1999;   99WO-US31274.
PR   18-FEB-2000; 2000WO-US04341.
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   21-MAR-2000; 2000WO-US07532.
PR   22-MAY-2000; 2000WO-US14042.
PR   02-JUN-2000; 2000WO-US15264.
PR   24-AUG-2000; 2000WO-US23328.
PR   10-NOV-2000; 2000WO-US30873.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   14-MAY-1999;   99US-0311832.
PR   25-AUG-1999;   99US-0380137.
PR   25-AUG-1999;   99US-0380138.
PR   25-AUG-1999;   99US-0380139.
PR   25-AUG-1999;   99US-0380142.
PR   15-SEP-1999;   99US-0397342.
PR   18-OCT-1999;   99US-0403297.
PR   12-NOV-1999;   99US-0423844.
PR   22-AUG-2000; 2000US-0644848.
PR   18-SEP-2000; 2000US-0664610.
PR   18-SEP-2000; 2000US-0665350.
PR   08-NOV-2000; 2000US-0709238.
PR   20-DEC-2000; 2000US-0747259.
PR   22-MAR-2001; 2001US-0816744.
PR   10-MAY-2001; 2001US-0854208.
PR   10-MAY-2001; 2001US-0854280.
PR   30-MAY-2001; 2001US-0870574.
PR   05-JUN-2001; 2001US-0874503.
PR   29-JUN-2001; 2001US-0869599.
PR   18-JUL-2001; 2001US-0908827.
PR   06-DEC-2001; 2001US-0006867.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-330485/31.
DR   N-PSDB; ACA58859.
XX
PT   New isolated antibody specifically binding a PRO polypeptide, useful
PT   for the preparation of a medicament for treating disorders with the
PT   aberrant expression or activity of the PRO polypeptide, such as tumor
PT   conditions and cancer  -
XX
PS   Disclosure; Page 169-171; 406pp; English.
XX
CC   The invention relates to an antibody that binds to a polypeptide with a
CC   fully defined sequence given in the specification. The methods and
CC   compositions (containing antibodies that specifically bind a PRO
CC   polypeptide) of the present invention are useful for the preparation of a
CC   medicament for the treatment of disorders associated with the aberrant
CC   expression or activity of the PRO polypeptide, such as tumour conditions
CC   and cancer. They can also be used to generate transgenic or knockout
CC   animals useful in the development and screening of therapeutically useful
CC   reagents. The PRO polypeptides and encoding nucleic acids can be used as
CC   molecular weight markers for protein electrophoresis, chromosome
CC   identification and tissue typing. The PRO polypeptides are useful to
CC   induce angiogenesis e.g wound healing; in the treatment of sports-related
CC   joint problems, articular cartilage defects, osteoarthritis or rheumatoid
CC   arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The
CC   antibodies may be used in various diagnostic, competitive binding and/or
CC   immunoprecipitation assays. The present sequence represents the amino
CC   acid sequence of a PRO polypeptide of the invention.
XX
SQ   Sequence   596 AA;

  Query Match             5.2%;  Score 148.5;  DB 24;  Length 596;
  Best Local Similarity   17.7%;  Pred. No. 0.0081;
  Matches  102;  Conservative   84;  Mismatches  255;  Indels  135;  Gaps   19;

Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||         ::|  :    |  : |:|              | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::       |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :   :|     :: |     |    |      :|    :  |     |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
                : :  | ||  :    :   ::|||   | :::||::|                : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
                : :  | : | |      :||   :: ::| | :  : |         | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
                |         : : ||         | |    ::|                    |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSARTVEN 410
                |    ||||     | :     | | |  :     :  :   | : ::| :: | | |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
                :: | :  | |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                  ||  :    || :||    |  :   |: | |
Db        402 -------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
                :| |:       :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477


RESULT 15
ABU72001
ID   ABU72001 standard; Protein; 596 AA.
XX
AC   ABU72001;
XX
DT   11-JUN-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO1342.
```

XX
KW Human; secreted and transmembrane polypeptide;
KW chromosome mapping; gene mapping; transgenic animal; knockout animal;
KW therapeutic agent screening; chromosome identification; tissue typing;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN US2003018183-A1.
XX
PD 23-JAN-2003.
XX
PF 01-MAY-2002; 2002US-0063512.
XX
PR 06-DEC-2001; 2001US-0006867.
XX
PA (GETH) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2003-330984/31.
DR N-PSDB; ACA60412.
XX
PT New secreted and transmembrane PRO polypeptides and nucleic acid
PT molecules encoding the polypeptides, useful in gene therapy or
PT preparing a medicament for treating a condition that is responsive to
PT the PRO polypeptide or antibody -
XX
PS Disclosure; Fig 100; 409pp; English.
XX
CC The invention describes novel isolated PRO polypeptides. The PRO
CC polypeptides or anti-PRO antibodies are useful in preparing a medicament
CC for treating a condition that is responsive to the PRO polypeptide or
CC antibody. The PRO nucleotide sequences may be used as hybridisation
CC probes in chromosome and gene mapping, or in generating antisense RNA
CC and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,
CC in assays to identify other proteins or molecules involved in binding
CC reaction, to generate transgenic animals or knockout animals, which in
CC turn are useful in the development and screening of therapeutically
CC useful reagents, for chromosome identification, and tissue typing. The
CC PRO polypeptides and nucleic acid molecules are also useful in gene
CC therapy, and as molecular weight markers for protein electrophoresis
CC purposes. The anti-PRO antibodies may be used in diagnostic assays for
CC PRO, or for the affinity purification of PRO from recombinant cell
CC culture or natural sources. This is the amino acid sequence of a novel
CC human secreted and transmembrane PRO poypeptide.
XX
SQ Sequence 596 AA;

Query Match 5.2%; Score 148.5; DB 24; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0081;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

```
Qy          4 GNLGHNPNVNNSIPPAPPLPSQTDGAGGRGQLINSTGPLGSRALFTPVRNSMADSGDNRA 63
              | | |     ||        ::|  :   |  : |:|               | :  |
Db         13 GLLLHLEAATNS--------NETSTSANTGSSVISSG--------------ASTATNSG 49

Qy         64 SDVPGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQIGSSV-FRVETQEDGKHIAVG 122
              | |    |:   :: | :| | |  ::  :    | :  | ::      |   |  ||
Db         50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSTASSGISIATN 107

Qy        123 QRNGVETS--VVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRIL 180
                :  :|    :: |     |    |      :|    :  |      |: :    |
Db        108 SESSTTSSGASTATNSE----SSTPSSGASTVTNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy        181 ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRSDPKLWLALGTVAT 240
              : :  | ||  :   :  ::|||   | :::||::|            : ||
Db        164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSTTSSG----------ASTAT 211

Qy        241 GLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEAFQNPDNQKVNIDE 300
              : :  | : | |      :||   :: ::| | :  : |        | : :
Db        212 N----SESSTVSSRASTATNSESSTT---SSGASTATNSESRTTSNGAGTATNSESSTTS 264

Qy        301 LG---------NAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQQAKR 351
               |          : : ||         | |    ::|                    |
Db        265 SGASTATNSDSSTVSSGA---STATNSESSTTSSGAST------------------ATN 302

Qy        352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSARTVEN 410
              |    ||||      | :      | | |  :      :  :    | : ::| :: | |  |
Db        303 SESSTTSSGASTATNSDSSTTSSGAGTATNSESSTVSSGISTVTNSESSTPSSGANTATN 362

Qy        411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGGPCRIRMLMLKHRCMIRR 470
              :: | :  |  |      |: |  |:  :: |:   |:| |
Db        363 SESSTTSSGANTATNSESSTVSSGASTATNSESS---TTSSG----------------- 401

Qy        471 CRLLILIRLFRIWGIQISVVYSTIQHPPRDTTDNGARLLGNPSAGIQSTYARLA------ 524
                                  ||  :    || :||    |  :   |: |  |
Db        402 --------------------VSTATNSESSTTSSGASTATNSDSSTTSSEASTATNSESS 441

Qy        525 -LSGGL----RHDMGGLTGGSNSAVNTSNNPPAPGS 555
               :| |:      :    : |:|:| |: ::  : ||
Db        442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477
```

Search completed: December 16, 2003, 08:52:56
Job time : 56.9828 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 48.5578 Seconds
(without alignments)
2917.574 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2795	99.8	549	2	O50190	O50190 escherichia
2	2676.5	95.6	550	2	O52147	O52147 escherichia
3	2496.5	89.2	552	2	Q9KWH9	Q9kwh9 escherichia
4	2206	78.8	551	2	O68258	O68258 escherichia
5	2077	74.2	547	2	Q9WXK1	Q9wxk1 escherichia
6	2076	74.1	547	2	Q9ETI1	Q9eti1 citrobacter
7	1781.5	63.6	538	2	Q47016	Q47016 escherichia
8	1781.5	63.6	538	2	Q47014	Q47014 escherichia
9	1763.5	63.0	538	2	O85508	O85508 escherichia
10	1528.5	54.6	558	16	Q9R396	Q9r396 escherichia
11	1516.5	54.2	558	2	O85506	O85506 escherichia
12	168.5	6.0	1158	5	O97169	O97169 drosophila
13	160.5	5.7	818	6	Q9N1P0	Q9n1p0 bos taurus
14	159.5	5.7	1265	2	Q9FDA0	Q9fda0 xanthomonas
15	156.5	5.6	1323	5	Q9VI63	Q9vi63 drosophila
16	156.5	5.6	1376	5	Q9BMZ9	Q9bmz9 drosophila
17	155.5	5.6	1321	5	Q962D1	Q962d1 drosophila
18	153.5	5.5	1015	5	Q8SZW9	Q8szw9 drosophila
19	152	5.4	1203	5	Q9N5K0	Q9n5k0 caenorhabdi
20	151.5	5.4	2793	16	Q8X2Q2	Q8x2q2 escherichia
21	151.5	5.4	2806	2	Q9KXA6	Q9kxa6 escherichia
22	151.5	5.4	2806	9	Q9T1K9	Q9t1k9 bacteriopha
23	151	5.4	934	5	Q9VN59	Q9vn59 drosophila
24	150.5	5.4	764	5	Q9BIT1	Q9bit1 plectreurys
25	149.5	5.3	2806	9	Q9XJM1	Q9xjm1 bacteriopha
26	149.5	5.3	2806	16	Q8X470	Q8x470 escherichia
27	149	5.3	832	2	O54356	O54356 moraxella c
28	149	5.3	2232	5	Q8IFX6	Q8ifx6 caenorhabdi
29	146	5.2	705	5	Q9GYB2	Q9gyb2 leishmania
30	146	5.2	1114	16	Q8PGR8	Q8pgr8 xanthomonas
31	146	5.2	1763	5	Q8I110	Q8i110 caenorhabdi
32	146	5.2	1829	5	Q22248	Q22248 caenorhabdi
33	145.5	5.2	1331	5	Q8MMB1	Q8mmb1 drosophila
34	145.5	5.2	3080	5	Q9V602	Q9v602 drosophila
35	145.5	5.2	3109	5	Q9BMQ0	Q9bmq0 drosophila
36	144.5	5.2	1296	5	O96506	O96506 drosophila
37	144.5	5.2	1383	5	Q9VCP6	Q9vcp6 drosophila
38	144	5.1	1589	6	O62672	O62672 bos taurus
39	143.5	5.1	673	11	Q8K1Q4	Q8k1q4 rattus norv
40	143.5	5.1	1283	5	Q8T5H0	Q8t5h0 anopheles g
41	143.5	5.1	1778	5	Q9NE65	Q9ne65 leishmania
42	143.5	5.1	2551	16	Q8CYI8	Q8cyi8 streptococc
43	142	5.1	3443	11	Q8JZM8	Q8jzm8 mus musculu
44	141.5	5.1	647	3	O14273	O14273 schizosacch
45	141.5	5.1	1488	16	Q8NPK0	Q8npk0 corynebacte

ALIGNMENTS

RESULT 1
O50190
ID O50190 PRELIMINARY; PRT; 549 AA.
AC O50190;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E2348/69;
RX MEDLINE=98050926; PubMed=9390560;
RA Kenny B., DeVinney R., Stein M., Reinscheid D.J., Frey E.A.,
RA Finlay B.B.;
RT "Enteropathogenic E. coli (EPEC) transfers its receptor for intimate
RT adherence into mammalian cells.";
RL Cell 91:511-520(1997).
DR EMBL; AF013122; AAB88410.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
KW Receptor.
SQ SEQUENCE 549 AA; 56843 MW; 40C8B8B234409A08 CRC64;

Query Match 99.8%; Score 2795; DB 2; Length 549;
Best Local Similarity 99.6%; Pred. No. 1.1e-148;
Matches 547; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTK 180

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420

Qy        421 PQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP 480

Qy        481 GYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAA 540

Qy        541 PTPGPVRFV 549
              |||||||||
Db        541 PTPGPVRFV 549
```

RESULT 2
O52147
ID O52147 PRELIMINARY; PRT; 550 AA.
AC O52147;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tir.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98254123; PubMed=9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
RT from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
DR EMBL; AF022236; AAC38390.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
SQ SEQUENCE 550 AA; 56509 MW; 19DD08A9BE9251CB CRC64;

Query Match 95.6%; Score 2676.5; DB 2; Length 550;
Best Local Similarity 96.0%; Pred. No. 4.9e-142;
Matches 531; Conservative 4; Mismatches 11; Indels 7; Gaps 2;

```
Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTK 180

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   : |
Db        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGG---NTPAQG 417

Qy        421 PQERRRFSD----RRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEV 476
              : | |    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 GTDATRAEDASLNRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEV 477

Qy        477 AADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSS 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db        478 AADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGESAVSS 537

Qy        537 VNAAPTPGPVRFV 549
              |||||||||||||
Db        538 VNAAPTPGPVRFV 550
```

RESULT 3
Q9KWH9
ID Q9KWH9 PRELIMINARY; PRT; 552 AA.
AC Q9KWH9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tir.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK01;
RA Abe A., Nagano H.;
RT "Analyses of type III secreted proteins and Tir in enteropathogenic
RT Escherichia coli O157:H45.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB036053; BAA96815.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
SQ SEQUENCE 552 AA; 57005 MW; A8D79EE22EE50A4B CRC64;

Query Match 89.2%; Score 2496.5; DB 2; Length 552;
Best Local Similarity 90.3%; Pred. No. 6e-132;
Matches 501; Conservative 11; Mismatches 34; Indels 9; Gaps 3;

```
Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||||||| ||||||||||||||||||||||||||||||||||||||:|||
Db          1 MPIGNLGNNVNSNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSVADS 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              |||||||||| |||||||||||||||||||||||||| || |||||||||| ||||||
Db         61 VDSRDIPGLPANPSRLAAATSETCLLGGFEVLHDKGPLDTLNQQIGPSAFRVEAQADGTH 120

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
```

```
                    ||||||||||||| |||||  |||||| |||||||||||||||||||||||||||||||| :
Db        121 AAIGEKNGLEVSVALSPQELQSLQSIDIEGKNRFVFTGGRGGSGHPMVTVASDIAEARMK 180

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              |||||||||||| |||||||||||||||||||||||||||||||||||:||||||||
Db        301 IDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQYARRQEELQL 360

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              |||||||||||| |||||||||||||||||||||||||||||||||||||   :  |
Db        361 SSGIGYGLSSALIGAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGG---NTPAQG 417

Qy        421 PQERRRFSD----RRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEV 476
                :  |  |    ||||||||||||||||||||||||||||  |||||  | ||||||||
Db        418 GTDATRAEDASLNRRDSQGSVASTHWSDSSSEVVNPYAEVGEPRNSLSTRQQEEHIYDEV 477

Qy        477 AADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAV-- 534
              |||| |||||||| : ||||||:|:||||||||||||||| ||||||||||  ||:||
Db        478 AADPVYSVIQNFSRNAPVTGRLMGSPGQGIQSTYALLANSAGLRLGMGGLTGSGESAVNT 537

Qy        535 SSVNAAPTPGPVRFV 549
              :: ||||||||||||
Db        538 ANANAAPTPGPVRFV 552


RESULT 4
O68258
ID   O68258      PRELIMINARY;      PRT;   551 AA.
AC   O68258;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98187918; PubMed=9529069;
RA   Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT   "Molecular analysis of Shiga toxigenic Escherichia coli O111:H-
RT   proteins which react with sera from patients with hemolytic-uremic
RT   syndrome.";
RL   Infect. Immun. 66:1467-1472(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
DR   EMBL; AF025311; AAC69249.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   551 AA;  56975 MW;  6EC95F76EF0F44CC CRC64;

  Query Match             78.8%;  Score 2206;  DB 2;  Length 551;
  Best Local Similarity   79.8%;  Pred. No. 1.1e-115;
  Matches  442;  Conservative   39;  Mismatches   65;  Indels    8;  Gaps    3;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

                    |||||||||| |:|||||||||||||:||| | ||:|||||||| ||||||||:|||
Db          1 MPIGNLGNNVNSNNLIPPAPPLPSQTDGASRGGAGQLINSTGALGSRLLFSPLRNSIADS 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              |||||||||| :|||| |||| |||||||||||||||| || ||| ||||:| |:||::
Db         61 VDSRDIPGLPVHPSRLATATSEICLLGGFEVLHDKGPLDTLNKQIGASAFRIEQQSDGSY 120

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
              ||||||||:|||| |: ||  |||:|| | | |||||||||| || ||| |||||||| :
Db        121 AAIGEKNGVEVSVILNSQELQSLQAIDIEDKGRFVFTGGRGGGGHSMVTPASDIAEARAK 180

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              |||||||:|||| | ::||||||||||||:|| ||||| ||:| ||||||||||||:||
Db        181 ILAKLDPNNHGGSQARNVDTRSVGVGSASGMDDSVVSETRTSSTASSVRSDPKFWVSIGA 240

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              |||||||||||| |||||||||||||||:|||:||||||:|||||||||||||||||:
Db        241 IAAGLAGLAATGITQALALTPEPDDPTTTDPEQAASAAESATRDQLTQEAFKNPENQKVS 300

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              ||  ||:|||||| ||:| :| :|||||| |||||||||||||||: |:|||||||:|
Db        301 IDEIGNSIPSGELKDDVVAKIEEQAKEAGEAARQQAVESNAQAQQRYDTQYARRQEELEL 360

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              |||||| ||||||| |||||||||||||||||||||||||||||||||||||   :  |
Db        361 SSGIGYSLSSALIVGGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGG---NTPAQG 417

Qy        421 PQERRRFSD----RRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEV 476
                :  |  |    ||||| | ||||||||:|| ||||||||| ||||  | | ||||||||
Db        418 GTDAIRAEDTSLNRRDSQRSTASTHWSDTSSAVVNPYAEVGEARNSSPARQAEEHIYDEV 477

Qy        477 AADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLA-NSGGLRLGMGGLTSGGETAVS 535
              |||| |||||||||:  |||||:||||||||||||:|  || |||||||||  | :||:
Db        478 AADPNYSVIQNFSGNNQVTGRLMGTPGQGIQSTYAILTNNSAGLRLGMGGLTGSGGSAVN 537

Qy        536 SVNAAPTPGPVRFV 549
              : |||||||| |||
Db        538 TANAAPTPGPGRFV 551


RESULT 5
Q9WXK1
ID   Q9WXK1      PRELIMINARY;      PRT;   547 AA.
AC   Q9WXK1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor (Translocated intimin receptor
DE   Tir).
GN   TIR.
OS   Escherichia coli, and
OS   Citrobacter rodentium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 67825;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=MPEC;
RA   Okutani A., Itoh K., Sasakawa C.;
RT   "Translocated intimin receptor(Tir) of murine pathogenic Escherichia
RT   coli O115a,c:K(B).";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=C.rodentium; STRAIN=DBS100;
RX   MEDLINE=21437640; PubMed=11553577;
RA   Deng W., Li Y., Vallance B.A., Finlay B.B.;
RT   "Locus of Enterocyte Effacement from Citrobacter rodentium: Sequence
RT   Analysis and Evidence for Horizontal Transfer among Attaching and
RT   Effacing Pathogens.";
RL   Infect. Immun. 69:6323-6335(2001).
DR   EMBL; AB026719; BAA77400.1; -.
DR   EMBL; AF311901; AAL06376.1; -.
```

```
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   547 AA;  56283 MW;  02CAC6D625FA6EE1 CRC64;

  Query Match             74.2%;  Score 2077;  DB 2;  Length 547;
  Best Local Similarity   76.0%;  Pred. No. 1.8e-108;
  Matches  418;  Conservative   44;  Mismatches   84;  Indels    4;  Gaps    3;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||||   |:|||||||||||||||| ||    ||||||:|||| ||||||:|: |:
Db          1 MPIGNLGNNNISNNLIPPAPPLPSQTDGATRGNGSSLISSTGSLGSRLLFSPLRSSIVDT 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              |||||:|||| :| |   ||||||| |||||||||||| || :|| | |||| | ||||
Db         61 VDSRDVPGLPEHPLRF--ATSETCLHGGFEVLHDKGPLDTLNQKIGSSVFRVEQQPDGTH 118

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
              |||| |:|:|||||:  |  ||||:|||| |||||||||||| ||||||||::|| :
Db        119 AAIGVKDGVEVSVTLNSSELQSLQSLDTEGNGRFVFTGGRGGSGHAMVTVASDISQAREK 178

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              |:||||||||||||:|||||||||||: |||||||||||| ||||||||||||||||||
Db        179 IIAKLDPDNHGGRQPKDIDTRSVGVGSASGMGDGVVSETHTSTTTSSVRSDPKFWVSVGA 238

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              |||||||||||| ||:|||| ||||||||||:||||||:|||||||:|||:||:|||||
Db        239 IAAGLAGLAATGIVQAVALTPAPDDPTTTDPDEAANAAEAATKDQLTKEAFQNPDNQKVN 298

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              ||  |||||||||  ||:|  ||| ||| ||| ||||||||||||||::|| |:||:|| |
Db        299 IDELGNAIPSGELKDDVVAQIADQAKVAGEQARQQAVESNAQAQQRHDDQQAKRQQELDL 358

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              |||||||||||||| ||||||||  ||||| | || |  |||:|:||||||  : :
Db        359 SSGIGYGLSSALIVGGGIGAGVTAMLHRRNPPTEQ-TIATTHSVIQQQTGGNTRAQGGAD 417

Qy        421 PQERRRFS-DRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAAD 479
                    |  ||||| ||||| |||:| :|||||||   :||: |   ||| ||||||
Db        418 TTGVENASLTRRDSQASVASTQWSDTSGDVVNPYAEGWMSRNNPSLLAPEEPIYDEVAPD 477

Qy        480 PGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNA 539
              | |||||:|||: ||||||:|:|||||||||||||:|||||||||| |||:| |: ||
Db        478 PNYSVIQHFSGNNPVTGRLVGSPGQGIQSTYALLASSGGLRLGMGGLTGGGESAGSAANA 537

Qy        540 APTPGPVRFV 549
              | |||  |||
Db        538 ATTPGVERFV 547


RESULT 6
Q9ETI1
ID   Q9ETI1      PRELIMINARY;      PRT;   547 AA.
AC   Q9ETI1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor Tir.
OS   Citrobacter rodentium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Citrobacter.
OX   NCBI_TaxID=67825;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1843-73T, and DBS100;
RX   MEDLINE=20553330; PubMed=11101562;
RA   Luperchio S.A., Newman J.V., Dangler C.A., Schrenzel M.D.,
RA   Brenner D.J., Steigerwalt A.G., Schauer D.B.;
RT   "Citrobacter rodentium, the Causative Agent of Transmissible Murine
RT   Colonic Hyperplasia, Exhibits Clonality: Synonymy of C. rodentium and
RT   Mouse-Pathogenic Escherichia coli.";
RL   J. Clin. Microbiol. 38:4343-4350(2000).

DR   EMBL; AF301618; AAG40758.1; -.
DR   EMBL; AF301617; AAG25642.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   547 AA;  56270 MW;  CB8318B301049C37 CRC64;

  Query Match             74.1%;  Score 2076;  DB 2;  Length 547;
  Best Local Similarity   76.0%;  Pred. No. 2.1e-108;
  Matches  418;  Conservative   44;  Mismatches   84;  Indels    4;  Gaps    3;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||||   |:|||||||||||||||| ||    ||||||:|||| ||||||:|: |:
Db          1 MPIGNLGNNNISNNLIPPAPPLPSQTDGATRGNGSSLISSTGSLGSRLLFSPLRSSIVDT 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              |||||:|||| :| |   ||||||| |||||||||||| || :|| | |||| | ||||
Db         61 VDSRDVPGLPEHPLRF--ATSETCLHGGFEVLHDKGPLDTLNQKIGSSVFRVEQQPDGTH 118

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
              |||| |:|:|||||:  |  ||||:|||| |||||||||||| ||||||||::|| :
Db        119 AAIGVKDGVEVSVTLNSSELQSLQSLDTEGTGRFVFTGGRGGSGHAMVTVASDISQAREK 178

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              |:||||||||||||:|||||||||||: |||||||||||| ||||||||||||||||||
Db        179 IIAKLDPDNHGGRQPKDIDTRSVGVGSASGMGDGVVSETHTSTTTSSVRSDPKFWVSVGA 238

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              |||||||||||| ||:|||| ||||||||||:||||||:|||||||:|||:||:|||||
Db        239 IAAGLAGLAATGIVQAVALTPAPDDPTTTDPDEAANAAEAATKDQLTKEAFQNPDNQKVN 298

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              ||  |||||||||  ||:|  ||| ||| ||| ||||||||||||||::|| |:||:|| |
Db        299 IDELGNAIPSGELKDDVVAQIADQAKVAGEQARQQAVESNAQAQQRHDDQQAKRQQELDL 358

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              |||||||||||||| ||||||||  ||||| | || |  |||:|:||||||  : :
Db        359 SSGIGYGLSSALIVGGGIGAGVTAMLHRRNPPTEQ-TIATTHSVIQQQTGGNTRAQGGAD 417

Qy        421 PQERRRFS-DRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAAD 479
                    |  ||||| ||||| |||:| :|||||||   :||: |   ||| ||||||
Db        418 TTGVENASLTRRDSQASVASTQWSDTSGDVVNPYAEGWMSRNNPSLLAPEEPIYDEVAPD 477

Qy        480 PGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNA 539
              | |||||:|||: ||||||:|:|||||||||||||:|||||||||| |||:| |: ||
Db        478 PNYSVIQHFSGNNPVTGRLVGSPGQGIQSTYALLASSGGLRLGMGGLTGGGESAGSAANA 537

Qy        540 APTPGPVRFV 549
              | |||  |||
Db        538 ATTPGVERFV 547


RESULT 7
Q47016
ID   Q47016      PRELIMINARY;      PRT;   538 AA.
AC   Q47016;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   TRANSLOCATED intimin receptor (Tir).
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REPEC 83/39;
RA   O'Brien R.A.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=REPEC 83/39;
RA   Krejany E.O.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RA   Agin T.S., Boedeker E.C.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 172-538 FROM N.A.
RC   STRAIN=REPEC RDEC-1;
RA   Krejany E.O.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-317 FROM N.A.
RC   STRAIN=RDEC-1;
RX   MEDLINE=97055784; PubMed=8900070;
RA   Agin T.S., Cantey J.R., Boedeker E.C., Wolf M.K.;
RT   "Characterization of the eaeA gene from rabbit enteropathogenic
RT   Escherichia coli strain RDEC-1 and comparison to other eaeA genes from
RT   bacteria that cause attaching-effacing lesions.";
RL   FEMS Microbiol. Lett. 144:249-258(1996).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RX   MEDLINE=98254123; PubMed=9593291;
RA   Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA   Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT   "The complete sequence of the locus of enterocyte effacement (LEE)
RT   from enteropathogenic Escherichia coli E2348/69.";
RL   Mol. Microbiol. 28:1-4(1998).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RX   MEDLINE=21153569; PubMed=11254564;
RA   Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,
RA   Boedeker E.C.;
RT   "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte
RT   Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1.";
RL   Infect. Immun. 69:2107-2115(2001).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RDEC-1;
RA   Boedeker E.C., Zhu C., Elliott S.J., Tonia T.S., Johnson L.A.,
RA   Thate T.E., Kaper J.B.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=83/39;
RA   Tauschek M., Strugnell R.A., Robins-Browne R.M.;
RT   "Characterization of the LEE pathogenicity islands of rabbit
RT   enteropathogenic Escherichia coli.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U59504; AAD19750.1; -.
DR   EMBL; AF045568; AAC15683.1; -.
DR   EMBL; U59503; AAB02941.1; -.
DR   EMBL; AF200363; AAK26722.1; -.
DR   EMBL; AF453441; AAL57549.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   538 AA;  55410 MW;  3132A969B7B3D06C CRC64;

  Query Match             63.6%;  Score 1781.5;  DB 2;  Length 538;
  Best Local Similarity   65.4%;  Pred. No. 6.3e-92;
  Matches  367;  Conservative   66;  Mismatches   93;  Indels   35;  Gaps   11;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| |   |||||||||||||||  |    ||:| | :||| ||:|:|||:||:
Db          1 MPIGNLGHNSNVRALIPPAPPLPSQTDGAG-GARNQLINSNGPMGSRLLFTPIRNSVADA 59

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
                ||| |||||||| | ||  ||  | |  ||||||| || ||: || | |||| : ||
Db         60 ADSRASDIPGLPTNPLRFAA--SEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDG 117

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
              :| |||:|||| :| || ||:||||:| ||||:||||||||:|| ||||||||||
Db        118 SHVAIGQKNGLETTVVLSEQEFSSLQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEAR 177

Qy        179 TRILAKLDP-DNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVS 237
              ||: ||:| |    ::| | ::            :| : | |||| ||:|:||| |:|
Db        178 QRIIDKLEPKDTKETKEPGDPNS----------GEGKIIEIHTSTSTSSLRADPKLWLS 226

Qy        238 VGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQ 297
              :| ||||| |:|||||||:|||||||| ||||| ||| ||:| |||||:|||:||:||
Db        227 LGTIAAGLIGMAATGIAQAVALTPEPDDPITTDPDAAANTAEAAAKDQLTKEAFQNPDNQ 286

Qy        298 KVNIDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEE 357
              ||||| |||||||||| ||:| |||:||| ||| |||:|:|||:||||:|:|:|||:|::|
Db        287 KVNIDENGNAIPSGELKDDVVAQIAEQAKAAGEQARQEAIESNSQAQQKYDEQHAKREQE 346

Qy        358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKV 417
              : ||||:|||:| |||: |||||||| ||||:|||||  | || |||  |    | :
Db        347 MSLSSGVGYGISGALILGGGIGAGVTAALHRKNQPAEQ--TITTRTVVDNQ----PTNNA 400

Qy        418 ALM-------PQERRRFSDRRDSQGSVASTHWSDSSS--EVVNPYAEVGGARNSLSAHQP 468
              :       |:|    : ||:|  |:|| : ||:||   | ||||:||  ||   |
Db        401 SAQGNTDTSGPEESP--ASRRNSNASLAS-NGSDTSSTGTVENPYADVGMPRNDSLARIS 457

Qy        469 EEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTS 528
              || ||||||||| |||||:|||: |||||:|||||||||||||||:||||||||||
Db        458 EEPIYDEVAADPNYSVIQHFSGNSPVTGRLVGTPGQGIQSTYALLASSGGLRLGMGGLTG 517

Qy        529 GGETAVSSVNAAPTPGPVRFV 549
              |||:|||: |||||||| |||
Db        518 GGESAVSTANAAPTPGPARFV 538


RESULT 8
Q47014
ID   Q47014      PRELIMINARY;      PRT;   538 AA.
AC   Q47014;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   TRANSLOCATED intimin receptor TIR (EspE protein).
GN   TIR OR ESPE.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REPEC 84/110/1, and E65/56;
RA   Krejany E.O.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=413/89-1;
RX   MEDLINE=98294040; PubMed=9632251;
RA   Deibel C., Kraemer S., Chakraborty T., Ebel F.;
RT   "EspE, a novel secreted protein of attaching and effacing bacteria, is
RT   directly translocated into infected host cells, where it appears as a
RT   tyrosine-phosphorylated 90 kDa protein.";
RL   Mol. Microbiol. 28:463-474(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=95ZG1;
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of Shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
```

```
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10;
RA   Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
RA   Boullier S., De Rycke J., Milon A., Oswald E.;
RT   "Role of Tir and Intimin in the pathogenesis of rabbit
RT   enteropathogenic Escherichia coli.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=413/89-1;
RA   Benkel P., Chakraborty T.;
RT   "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
RT   producing Escherichia coli.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U59502; AAC32028.2; -.
DR   EMBL; AJ223063; CAA11065.1; -.
DR   EMBL; AF070068; AAC69316.1; -.
DR   EMBL; AF132728; AAD27868.1; -.
DR   EMBL; AF113597; AAF03080.1; -.
DR   EMBL; AJ277443; CAC81869.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Hypothetical protein; Receptor.
SQ   SEQUENCE   538 AA;  55420 MW;  31D7A8E227B3D06C CRC64;

  Query Match           63.6%;  Score 1781.5;  DB 2;  Length 538;
  Best Local Similarity   65.4%;  Pred. No. 6.3e-92;
  Matches  367;  Conservative   66;  Mismatches   93;  Indels   35;  Gaps   11;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| |   |||||||||||||||  |    ||:| | :||| ||:|:|||:||:
Db          1 MPIGNLGHNPNVRALIPPAPPLPSQTDGAG-GARNQLINSNGPMGSRLLFTPIRNSVADA 59

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
               |||  |||||||| | ||  ||  | |  ||||||| || ||: || | |||| : ||
Db         60 ADSRASDIPGLPTNPLRFAA--SEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDG 117

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
              :| |||:|||| :| || ||:|||||:| ||||:|||||||:|| ||||||||||||
Db        118 SHVAIGQKNGLETTVVLSEQEFSSLQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEAR 177

Qy        179 TRILAKLDP-DNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVS 237
               ||: ||:| |    ::| | ::          :| : | ||||: ||:|:||| |:|
Db        178 QRIIDKLEPKDTKETKEPGDPNS----------GEGKIIEIHTSTSTSSLRADPKLWLS 226

Qy        238 VGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQ 297
              :| ||||| |:||||||||:|||||||| ||||| ||| ||:| |||||:|||:||:||
Db        227 LGTIAAGLIGMAATGIAQAVALTPEPDDPITTDPDAAANTAEAAAKDQLTKEAFQNPDNQ 286

Qy        298 KVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEE 357
              ||||| |||||||||| ||:| |||:||| ||| |||:|:||||:||||:|::||||:|::|
Db        287 KVNIDENGNAIPSGELKDDVVAQIAEQAKAAGEQARQEAIESNSQAQQKYDEQHAKREQE 346

Qy        358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKV 417
              : ||||:|||:| |||: |||||||| ||||:||||||  | || |||  |    | :
Db        347 MSLSSGVGYGISGALILGGGIGAGVTAALHRKNQPAEQ--TITTRTVVDNQ----PTNNA 400

Qy        418 ALM-------PQERRRFSDRRDSQGSVASTHWSDSSS--EVVNPYAEVGGARNSLSAHQP 468
              :         |:|    : ||:|  |:|| : ||:||   | ||||:||  ||   |
Db        401 SAQGNTDTSGPEESP--ASRRNSNASLAS-NGSDTSSTGTVENPYADVGMPRNDSLARIS 457

Qy        469 EEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTS 528
              || ||||||||| |||||:|||: ||||||:||||||||||||||:||||||||||||
Db        458 EEPIYDEVAADPNYSVIQHFSGNSPVTGRLVGTPGQGIQSTYALLASSGGLRLGMGGLTG 517

Qy        529 GGETAVSSVNAAPTPGPVRFV 549
              |||:|||: |||||||| |||
Db        518 GGESAVSTANAAPTPGPARFV 538


RESULT 9
```

```
O85508
ID   O85508      PRELIMINARY;      PRT;   538 AA.
AC   O85508;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Translocated intimin receptor Tir.
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EPEC87A;
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
DR   EMBL; AF070069; AAC69318.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   538 AA;  55602 MW;  447052A0E3214D6D CRC64;

  Query Match           63.0%;  Score 1763.5;  DB 2;  Length 538;
  Best Local Similarity   64.5%;  Pred. No. 6.5e-91;
  Matches  362;  Conservative   71;  Mismatches   93;  Indels   35;  Gaps   11;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| |   |||||||||||||||  |    ||:| | :||| ||:|:|||:||:
Db          1 MPIGNLGHNPNVRALIPPAPPLPSQTDGAG-GARNQLINSNGPMGSRLLFTPIRNSVADA 59

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
               |||  |||||||| | ||  ||  | |  ||||||| || ||: || | |||| : ||
Db         60 ADSRASDIPGLPTNPLRFAA--SEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDG 117

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
              :| |||:|||| :| || ||:|||||:| ||||:|||||||:|| ||||||||||||
Db        118 SHVAIGQKNGLETTVVLSDQEFFSIQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEAR 177

Qy        179 TRILAKLDP-DNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVS 237
               ||: ||:| |    ::| | ::          :| : | ||||:  |:|:||| |:|
Db        178 QRIIDKLEPKDTKETKEPGDPNS----------GEGKIIEIHTSTSTFSLRADPKLWLS 226

Qy        238 VGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQ 297
              :| ||||| |:||||||||:|||||||||||||||| ||: ||:||||:||||||::|: |
Db        227 LGTIAAGLIGMAATGIAQAVALTPEPDDPTTTDPDTAASTAEAATKDRLTQEAFQDPDKQ 286

Qy        298 KVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEE 357
              ||||| |||||||||| ||:| |||:||| ||| |||:|:||||:||||:|::|||:|::|
Db        287 KVNIDENGNAIPSGELIDDVVAQIAEQAKAAGEQARQEAIESNSQAQKKYDEQHAKREQE 346

Qy        358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKV 417
              : ||||:|||:| |||: |||||||| ||||:||||||  | || |||  |    | :
Db        347 MSLSSGVGYGISGALILGGGIGAGVTAALHRKNQPAEQ--TITTRTVVDNQ----PTNNA 400

Qy        418 ALM-------PQERRRFSDRRDSQGSVASTHWSDSSS--EVVNPYAEVGGARNSLSAHQP 468
              :         |:|    : ||:|  |:|| : ||:||   | ||||:||  ||   |
Db        401 SAQGNTDTSGPEESP--ASRRNSNASLAS-NGSDTSSTGTVENPYADVGMPRNDSLARIP 457

Qy        469 EEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTS 528
              || ||||||||| |||||:|||: ||||||:||||||||||||||:||||||||||||
Db        458 EEPIYDEVAADPNYSVIQHFSGNSPVTGRLVGTPGQGIQSTYALLASSGGLRLGMGGLTG 517

Qy        529 GGETAVSSVNAAPTPGPVRFV 549
              |||:|||: ||:||||| |||
Db        518 GGESAVSTANASPTPGPARFV 538
```

```
RESULT 10
Q9R396
ID   Q9R396      PRELIMINARY;      PRT;   558 AA.
AC   Q9R396;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Translocated intimin receptor Tir (Putative translocated intimin
DE   receptor protein).
GN   TIR OR Z5112 OR ECS4561.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=86/24;
RX   MEDLINE=99242825; PubMed=10225900;
RA   DeVinney R., Stein M., Reinscheid D., Abe A., Ruschkowski S.,
RA   Finlay B.B.;
RT   "Enterohemorrhagic Escherichia coli O157:H7 produces Tir, which is
RT   translocated to the host cell membrane but is not tyrosine
RT   phosphorylated.";
RL   Infect. Immun. 67:2389-2398(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=EDL933;
RX   MEDLINE=98339885; PubMed=9673266;
RA   Perna N.T., Mayhew G.F., Posfai G., Elliott S., Donnenberg M.S.,
RA   Kaper J.B., Blattner F.R.;
RT   "Molecular evolution of a pathogenicity island from enterohemorrhagic
RT   Escherichia coli O157:H7.";
RL   Infect. Immun. 66:3810-3817(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AF125993; AAD29391.1; -.
DR   EMBL; AF071034; AAC31506.1; -.
DR   EMBL; AE005595; AAG58825.1; -.
DR   EMBL; AP002566; BAB37984.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor; Complete proteome.
SQ   SEQUENCE   558 AA;  58022 MW;  99C417222D4B4AA1 CRC64;

  Query Match             54.6%;  Score 1528.5;  DB 16;  Length 558;
  Best Local Similarity   56.2%;  Pred. No. 9.8e-78;
  Matches  327;  Conservative   60;  Mismatches  138;  Indels   57;  Gaps   11;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| | |: |||||||||||||  || | ||:||| ||||:||:|:|||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
               |:|  |:|||| || |||| || | |||||| |||| || ||| | |||| | ||
Db         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
              | |:|::||:| || || ||:: ||||| |||::|||||||||:|| |||||||| |||
Db        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176

Qy        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230
              |||  |:|   |           | | | : |:      : |    :|| |||: ||:||
Db        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227

Qy        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290
              ||| |:::| :| || ||||||| |||||||||| ||||||| ||:| |:||:||||:||
Db        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287

Qy        291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350
              |:||:|||||||  |||||||  | ||:|  | :||| ||| |:|||:|:|||||::|:::|
Db        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347

Qy        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403
              |:|||||::||| ||||| |||: ||||  || ||||:||| ||||||||         |
Db        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSART 407

Qy        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSS-SEVVNPYAEVGGARNS 462
              |  :      |         |         || |  | :|| :  ||   | ||||:|
Db        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGTVQNPYADV-----K 462

Qy        463 LSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVT---------------GRLIGTPGQGIQ 507
              | |         :|      : :||   :  |                 ||:| |  |||
Db        463 TSLHD------SQVPTSNSNTSVQNMGNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQ 516

Qy        508 STYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549
              |||| || |||||  ||||| |  :||:: |  | ||  |||
Db        517 STYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 558


RESULT 11
O85506
ID   O85506      PRELIMINARY;      PRT;   558 AA.
AC   O85506;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Translocated intimin receptor Tir.
GN   TIR.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=95SF2;
RX   MEDLINE=99003184; PubMed=9784578;
RA   Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT   "Translocated intimin receptors (Tir) of shiga-toxigenic Escherichia
RT   coli isolates belonging to serogroups O26, O111, and O157 react with
RT   sera from patients with hemolytic-uremic syndrome and exhibit marked
RT   sequence heterogeneity.";
RL   Infect. Immun. 66:5580-5586(1998).
DR   EMBL; AF070067; AAC69314.1; -.
DR   InterPro; IPR003536; Tir_receptor.
DR   Pfam; PF03549; Tir_receptor; 1.
DR   PRINTS; PR01370; TRNSINTIMINR.
KW   Receptor.
SQ   SEQUENCE   558 AA;  58175 MW;  CA2CDDAC94527C2E CRC64;

  Query Match             54.2%;  Score 1516.5;  DB 2;  Length 558;
  Best Local Similarity   55.8%;  Pred. No. 4.6e-77;
  Matches  325;  Conservative   59;  Mismatches  141;  Indels   57;  Gaps   11;
```

```
Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
Db         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
Db        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176

Qy        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230
Db        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227

Qy        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290
Db        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASETETATRDQLTKEA 287

Qy        291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350
Db        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347

Qy        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403
Db        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSART 407

Qy        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSS-SEVVNPYAEVGGARNS 462
Db        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTLSTFFDTSSIGTVQNPYADV-----K 462

Qy        463 LSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVT---------------GRLIGTPGQGIQ 507
Db        463 TSLHD------SQVPTSNSNTSVQNMGNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQ 516

Qy        508 STYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549
Db        517 STYARLALRGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 558
```

RESULT 12
O97169

```
ID   O97169      PRELIMINARY;      PRT;  1158 AA.
AC   O97169;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hairless protein.
GN   H OR HAIRLESS.
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Marquart J., Alexief-Damianof C., Preiss A., Maier D.;
RT   "Rapid divergence in the course of Drosophila evolution reveals
RT   structural important domains of the Notch antagonist Hairless.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ232783; CAB38221.1; -.
DR   FlyBase; FBgn0019190; Dhyd\H.
SQ   SEQUENCE   1158 AA;  119635 MW;  0B611D3739CF397F CRC64;

  Query Match             6.0%;  Score 168.5;  DB 5;  Length 1158;
  Best Local Similarity   18.9%;  Pred. No. 0.39;
  Matches  130;  Conservative   89;  Mismatches  238;  Indels  231;  Gaps   26;

Qy          6 LGNNVNGNHL--------------------IPPAPPLPSQTDGAARGGTGHLISSTGALG 45
Db         57 LKNGINGTNTAAAAAAATLLTAAAAAVTFQLPAAPELKPTTSATAANATNSTTSSSTATS 116

Qy         46 SRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLL-GGFEVLHDKGPLDILNTQ 104
Db        117 TSNGTTPALSSTKTNSSSNS-----SSSSLIMATASATALVAGGAGVTAPKAPADVMAGV 171

Qy        105 IGPSAFRVEVQADGTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGS- 163
Db        172 L--------------------------------DYSALNSPNATKSSKFVGTGNGSSSF 198

Qy        164 ---GHPMVTVASD-----------------IAEARTRILAKLDPDNHGG-----RQPKDV 198
Db        199 DMGRHPISMRSNNSMSGYGGRLQFFKDGKFILE-----LARAKDGDKGGWVSVPRKPFRT 253

Qy        199 DTRSVGVGSASGIDDGVVSETHT----------STTNSSVRSDP----KFWVSVGAIAAG 244
Db        254 PSAAT---SATVTPTSAVTTTYPKNENSTSLSFSDDNSSIQSSPWQRDQPWKQT------ 304

Qy        245 LAGLAATGIAQALAL---------------------TPEPDDPT-TTDPDQAANAAESAT 282
Db        305 ---RPRRGISKELSLYYQRPRYNVLSQSARQAAMRKRRRPYEPTIATENHQSIFERVSAQ 361

Qy        283 KDQLTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQ 342
Db        362 EN--GDETLKAVEVDTKAPEIQSDDIKSETIKEQLTEGKAKSDSKACSSDNKDLNDSSAK 419

Qy        343 AQQRYEDQHARRQEE-LQLSSGIGYGLSSALIVAGGIG---------AGVTTA----LHR 388
Db        420 AEATLEKMNTSEDEDAMSVTVDDGAPTTSDCLVNGNTSEVNGEPKAHTKCETASEKDTRR 479

Qy        389 RNQPAEQTTTTTT-------------HTVVQQQTGGIPQH-------------KVALMPQ 422
Db        480 TGTPATKTTTMMTDIRLKKQKPRAKLNSIIQKLIDGVPARLEQLSKTPAVAAATASAVAS 539

Qy        423 ERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHI---------Y 473
Db        540 TADRISSNSSGGGAIGSL--SHSLAHKVSPPSSAAAASRLVEYHHHHQHVSPRKRILREF 597

Qy        474 DEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTS--GGE 531
Db        598 EKVSLEDNNGCVNNGSGVG------------------------GGISGGAGGKRSRAKGS 633

Qy        532 TAVSSV-------NAAP-------TPGP 545
Db        634 TATSAVTTKSMPINLAPPQAKVLSTPTP 661
```

RESULT 13
Q9N1P0

```
ID   Q9N1P0      PRELIMINARY;      PRT;   818 AA.
AC   Q9N1P0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Submaxillary mucin (Fragment).
GN   BSM1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=breed Angus;
RX   MEDLINE=20223253; PubMed=10759843;
RA   Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT   "The central domain of bovine submaxillary mucin consists of over 50
RT   tandem repeats of 329 amino acids: chromosomal localization of the
RT   BSM1 gene and relations to ovine and porcine counterparts.";
RL   Eur. J. Biochem. 267:2208-2217(2000).
DR   EMBL; AF178428; AAF67279.1; -.
FT   NON_TER       1      1
FT   NON_TER     818    818
SQ   SEQUENCE   818 AA;  73014 MW;  B5BB44F84F66F86B CRC64;
```

```
  Query Match             5.7%;  Score 160.5;  DB 6;  Length 818;
  Best Local Similarity   21.6%;  Pred. No. 0.68;
  Matches  122;  Conservative   64;  Mismatches  221;  Indels  157;  Gaps   21;

Qy         10 VNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGL 69
              |:|  : | : |  |:    :  |:   :: :|| |: :  |  |::       |  ||
Db        174 VSGAAVSPGSSPGRSRATAVSGEGSQPTVALSGATGTSAGPSGTRST------SSAIPAT 227

Qy         70 PTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTHAAIGEKNGL 129
              | : :  ||                :|   |  |:| |      || |  || :
Db        228 PGSTTGRAAGAGTPA------------VDSQQTASLPAAARPTALGPGTSAPSGETSES 274

Qy        130 EVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTRI----LAKL 185
               ||          ::    |      |    :  | ||  : :  : ||:    :: |
Db        275 RSSVP------GGSETTQQPG------AGSESPTLSPGVTRTTALRGSETRVPSTGVSGL 322

Qy        186 DPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST---TNSSVRSDPKFWVSVGAIA 242
                ||          : | |: ||     |  || ||    ||  |    |    |: |: |
Db        323 PGSTQGG------SAATGGSGAGSGPTAPVSGETRTSVISGTNVPVSGAP---VTPGSSA 373

Qy        243 -------------------AGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATK 283
                              :| || :|||   ::   |    | | :|   :  | ||
Db        374 GSSGAPGTGGPGSETASPLSGAAGTSATGSGTSI---PPSGAPVTPEPPLISTGA-SAGP 429

Qy        284 DQLTQEAFKNPENQKVNIDANGNAIP-SGELXDDIVEQIAQQAKEAGEVARQQAVESNAQ 342
               ::      |     ::  :| ::| ||          :        | |
Db        430 PASSESTVTLPGATGTDVLRSGTSLPVSG-------GAVTPAPSPGGSSA---------- 472

Qy        343 AQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTH 402
                              ::| | | :: :  :|  || |
Db        473 ------------------TAGPGVGSATTVQASGATGADVL------------------- 495

Qy        403 TVVQQQTGGIPQHKVALMP--QERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGAR 460
                  :    :|   ||: |      |  :    ||||  :   | ::   | |      | |
Db        496 ----RSGTSLPVSGVAVSPGSSPGRSGATAVSSQGSQPTVALSGATGTSVGP----SGTR 547

Qy        461 NSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLR 520
               | ||          : | ||       | :|   |   ||||   | | :| | :
Db        548 FSSSA----------IPATPG-------STTGRAAG--AGTPGVDSQQTASLPAAARPTA 588

Qy        521 LGMGGLTSGGETAVSSVNAAPTPG 544
               || |     |||   | : :  ||
Db        589 LGPGTSAPSGET---SESRSSVPG 609


RESULT 14
Q9FDA0
ID   Q9FDA0      PRELIMINARY;      PRT;  1265 AA.
AC   Q9FDA0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative outer membrane protein XadA.
GN   XADA.
OS   Xanthomonas oryzae (pv. oryzae).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=64187;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ray S.K., Rajeshwari R., Sonti R.V.;
RT   "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
RT   that is involved in virulence.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF288222; AAG01335.1; -.
DR   InterPro; IPR005594; YadA.
DR   Pfam; PF03895; YadA; 1.
SQ   SEQUENCE   1265 AA;  119856 MW;  21762579B5EC70A2 CRC64;

  Query Match             5.7%;  Score 159.5;  DB 2;  Length 1265;
  Best Local Similarity   23.4%;  Pred. No. 1.4;
  Matches  136;  Conservative   73;  Mismatches  229;  Indels  143;  Gaps   29;

Qy          3 IGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISS-TGALGSRSLFSPLR------N 55
              :||    :  |      :  : |      | |     :|: | ||||::: | |:      |
Db        592 LGNTARALGDNSSAVGSNAVASDIGATANGAGAQALSAYTIALGSKAVASDLQAIAEGFN 651

Qy         56 SMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQ 115
              | | :| |  :  |   :  ||::|   : :    :                :| :|     | |
Db        652 STASNVGSTALGGFSQSSGRLSSALGYSAVASSVD----------STAVG-----VAAQ 695

Qy        116 ADGTHA-AIGE--KNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVAS 172
              | |  | ||||  |   | || :   :|          ||     |  |   :      :
Db        696 ATGVSAVAIGEISKATGEESVAVGGGAFSGWIPTQASGKGAAAFGAGAWATADYTTAIGR 755

Qy        173 D-IAEA--RTRILAKLD--PDN----HGGRQPKDVDTRSVGV-GSASGIDD-------GV 215
              |  |:    | :    |   ||     || : | |    :||  ||:||:         |
Db        756 DSYADGVNATAVGQSADALADNTLALGGGSRAKAVGASVIGVDASATGINSTGVGRQVNV 815

Qy        216 VSETHTST-TNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQA 274
              :  |    |   || ||        :|  :| | |   ||| | ::||       :| | |
Db        816 IGENAVSVGYNSYVRQS-----AVNGVALG-ANAGATG-ADSVAL---GSGSSTYDADTV 865

Qy        275 ANAAESATKDQLTQEAFKNPENQKVNI------DANGNAIPSGELXDDIVEQIAQQAKEA 328
              :  : :      |:          ||:       |: :||  |:| : :
Db        866 SVGSGNGRGGPATRRI--------VNVGAGAVASASTDAINGGQLFESL---------- 906

Qy        329 GEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH- 387
                          |||                  :| :| |   | | | |:    |  :
Db        907 ----------SNA------------------ASFLGGG---AAIGAQGVFVAPTYLIQG 934

Qy        388 --RRNQPAEQTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDS 445
                   |  |   :  | :  ::|| | :  |      |  |:|   :  ::|:|  |
Db        935 ASYNNVGAALTALDSKVTELDARSGGTPANTAA------RTASLRTATVPAMAATAVSAV 988

Qy        446 SSEV----VNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGT 501
              || |    ::  | | |          |   :   :      :|:   : :  |||  ||
Db        989 SSNVASTAIDATAGVQGT--------PTAAVVGSITPAAISTVVGTAAVANNVTGTAIG- 1039

Qy        502 PGQGIQSTYALLANSGGL----RLGMGGLTS-GGETAVSSV 537
                    | ||  ||   :    |:   | |: |   | :::|
Db       1040 -----GSAYAHGANDTAIGSNARVNADGSTAVGANTQIAAV 1075


RESULT 15
Q9VI63
ID   Q9VI63      PRELIMINARY;      PRT;  1323 AA.
AC   Q9VI63; Q9NHX4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG1070 protein (Alhambra).
GN   ALHAMBRA OR CG1070.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
```

```
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Perrin L., Dura J.M.;
RT   "Alhambra, a Drosophila homolog of mammalian AF-10 and AF-17.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003672; AAF54062.2; -.
DR   EMBL; AF217960; AAF72595.1; -.
DR   FlyBase; FBgn0037471; Alhambra.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   SMART; SM00249; PHD; 2.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
SQ   SEQUENCE   1323 AA;  132884 MW;  D53C0C8AF392F9A6 CRC64;

  Query Match             5.6%;  Score 156.5;  DB 5;  Length 1323;
  Best Local Similarity   20.2%;  Pred. No. 2.2;
  Matches  135;  Conservative   85;  Mismatches  270;  Indels  177;  Gaps   25;

Qy          3 IGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLIS-----STGALGSRSLFSP-LRNS 56
              |||: |::| |    |     : : |   ||:| :||     :| :  |:|  :|    |
Db        322 IGNISNSLNNN---LPGGSSSTSSAGNVPGGSGGIISASSGGATQSTSSQSSTAPGTTKS 378

Qy         57 MADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA 116
              | |  |         :   |: :||    |     |          |    ||     :
Db        379 SASSSSSSSNSYKEKHSKSLSKSTSSKDKDG-----KDSSTNSANNNFTNSSASSTSSNS 433

Qy        117 DGTHAAIGEKNGLEVS-------VTLSPQEWSSLQSIDTE-GKNRFVFTGGRGGSG-HPM 167
                |     ||:  ::|       |  :    |:  ||:|:   :    | | ||:| |
Db        434 SSTR----EKSSSKLSKNKDSNQVPSATSSLSTTSSINTQPSSSTSTATAGSGGTGTHVS 489

Qy        168 VTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGV------------ 215
              : || :  | :          |        : |   | |||:|    |
Db        490 SSAASGLNSAPS-------TTNEHSNHAHTLSTNGTGAGSAAGKLQSVSNLSNSSSGFGS 542

Qy        216 ----VSETHTSTTNSSV-------RSDPKFWVSVGAIAAGLAGLAA--TGIAQALALTPE 262
                || : :|| | |         |: :    |: |: : |  |  ||    : | .|
Db        543 DLRSVSTSSSSTVNDSTGGFGSNSNSERENLSGAGSSASNMPGTIAPGTGGVSSSAAT-- 600

Qy        263 PDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNID---ANGNAIPSGELXDDIVE 319
              : :|     ::: | | |    :  :  |  ::|   |   :  :   ||   :|
Db        601 --NLSTNKGGSSSSTANSLTSTSTSSGSSSNSSSKKRKADSAKSTSSISTSGSALEDNNS 658

Qy        320 QIAQ-QAKEAGEVARQQAVESNAQAQQRYEDQHARRQEEL-------QLSSGIGYGLSSA 371
               |::  |:       | |:      ::  |     |:: ||          :: :   |:|:
Db        659 LISRYDIKDV-----QVALTPLTDFEKEIEKSSKRQRTELSPPTHQTSATAEVNAPLASS 713

Qy        372 LIVAGGIGAGVTTALHRRNQPAEQTTT--------------------TTTHTVVQQQTG 410
              :  : |  | |      ||  |||                       |       :|
Db        714 TSASIAVTASATAA---SAPPATGTTTLGSSISGNAGSTSSGNSSGVSATGGAQSAVGSG 770

Qy        411 GIPQHKVALMPQERRRFSDRRDSQGSVASTHWSD---------------SSSEVVNPYAE 455
              | |: : :          |    |  : :: | |:                |||   : |
Db        771 GYPKTESSKSSGTASAGSGSGSSSNTSSTKHGSNIKDISSSSNQQASTASSSSAPSLYVS 830

Qy        456 V--------------------------GGARNSLSAHQPEEHIYDEVAADP-------- 480
              |                          :|:| :  | :: : : :   |
Db        831 VPLSTANVPGINLPTSSTSSSTTSESHSASSRSSGAQSQHQQQLSNALVGPPMGGSAGAF 890

Qy        481 --------GYSVIQNFSG-SGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGG--LTSG 529
                      ||||  || | |  | |:              ||||  :   |  | ||
Db        891 HGGTTSAGSSSVIQQQSGKSSPALGTLVS------------GNSGGSIISASGFPLPSG 937

Qy        530 GETAVSS 536
                || ::
Db        938 NLTATTT 944


Search completed: December 16, 2003, 08:55:42
Job time : 52.5578 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 21.306 Seconds
(without alignments)
1211.758 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	157.5	5.6	494	1	SOX9_CHICK	P48434 gallus gall
2	150.5	5.4	1608	1	HLYA_SERMA	P15320 serratia ma
3	150	5.4	1005	1	Y456_CHLTR	O84462 chlamydia t
4	144	5.1	883	1	E74B_DROME	P11536 drosophila
5	140	5.0	2132	1	PGCA_MOUSE	Q61282 mus musculu
6	137	4.9	1532	1	IGA_NEIGO	P09790 neisseria g
7	136	4.9	1569	1	YPJA_ECOLI	P52143 escherichia
8	135.5	4.8	2124	1	PGCA_RAT	P07897 rattus norv
9	134.5	4.8	1150	1	APMU_PIG	P12021 sus scrofa
10	132.5	4.7	1025	1	SLAP_CAUCR	P35828 caulobacter
11	132	4.7	666	1	FLID_VIBCH	Q9kq63 vibrio chol
12	132	4.7	827	1	CSG_HALVO	P25062 halobacteri
13	131.5	4.7	618	1	ZEST_DROVI	Q24762 drosophila
14	130.5	4.7	940	1	MAZ4_SCHCO	P37938 schizophyll
15	130	4.6	1260	1	ALS1_CANAL	P46590 candida alb
16	130	4.6	1596	1	MAM_DROME	P21519 drosophila
17	130	4.6	2703	1	NOTC_DROME	P07207 drosophila
18	129	4.6	1341	1	ACIN_HUMAN	Q9ukv3 homo sapien
19	129	4.6	2415	1	PGCA_HUMAN	P16112 homo sapien
20	128	4.6	827	1	XANP_XANS2	Q60106 xanthomonas
21	127.5	4.6	509	1	SOX9_PIG	O18896 sus scrofa
22	127	4.5	784	1	SP4_HUMAN	Q02446 homo sapien
23	126.5	4.5	2493	1	CYAA_USTMA	P49606 ustilago ma
24	125.5	4.5	673	1	Y552_HUMAN	O60299 homo sapien
25	125	4.5	2442	1	CBP_HUMAN	Q92793 homo sapien
26	124	4.4	977	1	BAB1_DROME	Q9w0k7 drosophila
27	124	4.4	1589	1	PHP_DROME	P39769 drosophila
28	123.5	4.4	785	1	SP1_HUMAN	P08047 homo sapien
29	123.5	4.4	829	1	E74A_DROME	P20105 drosophila
30	123.5	4.4	2464	1	MAPB_MOUSE	P14873 mus musculu
31	123.5	4.4	3591	1	FHAB_BORPE	P12255 bordetella
32	123	4.4	980	1	SYN_DROME	Q24546 drosophila
33	122.5	4.4	1227	1	JIP3_DROME	Q9gqf1 drosophila
34	122.5	4.4	1306	1	MSB2_YEAST	P32334 saccharomyc
35	122.5	4.4	2541	1	TLN1_MOUSE	P26039 mus musculu
36	121.5	4.3	1338	1	ACIN_MOUSE	Q9jix8 mus musculu
37	121.5	4.3	1783	1	RAA3_CHLRE	Q9fec4 chlamydomon
38	120.5	4.3	314	1	YK83_MYCTU	Q10691 mycobacteri
39	120.5	4.3	1469	1	BCL9_DROME	Q961d9 drosophila
40	120	4.3	706	1	HTF4_MOUSE	Q61286 mus musculu
41	120	4.3	995	1	YIQ9_YEAST	P40442 saccharomyc
42	120	4.3	1505	1	CUT1_HUMAN	P39880 homo sapien
43	120	4.3	2541	1	TLN1_HUMAN	Q9y490 homo sapien
44	119	4.2	1953	1	BIGA_SALTY	P25927 salmonella
45	119	4.2	2175	1	HMCU_DROME	P10180 drosophila

ALIGNMENTS

RESULT 1
SOX9_CHICK
ID SOX9_CHICK STANDARD; PRT; 494 AA.
AC P48434;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-9.
GN SOX9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Uwanogho D., Rex M., Cartwright E.J., Pearl G., Scotting P.J.,
RA Sharpe P.T.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 1 HMG box domain.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U12533; AAB09663.1; -.
DR HSSP; P48436; 1SX9.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS50118; HMG_BOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 106 174 HMG BOX.
FT DOMAIN 215 220 POLY-SER.
FT DOMAIN 342 360 GLN/PRO-RICH.
FT DOMAIN 394 401 POLY-GLN.
SQ SEQUENCE 494 AA; 54942 MW; AF41D483B5016349 CRC64;

Query Match 5.6%; Score 157.5; DB 1; Length 494;
Best Local Similarity 18.9%; Pred. No. 0.031;
Matches 74; Conservative 48; Mismatches 142; Indels 127; Gaps 11;

```
Qy      103 TQIGPSAFRVEVQADGTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGG 162
            | | |:|   :|||                 |||  ||:  : : |::    :|   |
Db      197 THISPNAIFKALQAD-----------------SPQSSSSISEVHSPGEH----SGQSQG 234

Qy      163 SGHPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTS 222
                | |  :|  :   : | :  |  |||||  :| | | :|  |     |:|   |
Db      235 PPTPPTTPKTDAQQPGKQDLKREGPLAEGGRQPPHIDFRDVDIGELS---SDVISNIETF 291

Qy      223 TTNS-------------SVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTT 268
```

```
Db        292 DVNEFDQYLPPNGHPGVPATHGQVTTYSGTYGISSSASSPAGAGHAWMAKQQPQPPQPPA 351

Qy        269 TDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEA 328

Db        352 QPPAQ--------------------------------HTLPS-------TEREQGPAQQR 372

Qy        329 GEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIG------YGLSSALIVAGGIGAGV 382

Db        373 PHIKTEQLSPSHNSEQQQHPEQQQQQQQQL----GYGSFNLQHYGFSYPPI--------- 419

Qy        383 TTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVAS--T 440

Db        420 --------------TRSEYDYTEHQNSG-------------SYYSHAAGQSGSLYSTFT 451

Qy        441 HWSDSSSEVVNPYAEVGGARNSLSAHQPEEH 471

Db        452 YMNPTQRPMYTPIADTSGVPTIPQTHSPQQH 482


RESULT 2
HLYA_SERMA
ID   HLYA_SERMA     STANDARD;      PRT;  1608 AA.
AC   P15320;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   Hemolysin precursor.
GN   SHLA.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Serratia.
OX   NCBI_TaxID=615;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC   STRAIN=SN8;
RX   MEDLINE=88257037; PubMed=3290200;
RA   Poole K., Schiebel E., Braun V.;
RT   "Molecular characterization of the hemolysin determinant of Serratia
RT   marcescens.";
RL   J. Bacteriol. 170:3177-3188(1988).
CC   -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC       CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC       DEFINED.
CC   -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC       ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
CC       REQUIRES SHLB FUNCTION.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22618; AAA50323.1; -.
DR   PIR; A28182; A28182.
KW   Hemolysis; Toxin; Outer membrane; Signal.
FT   SIGNAL        1     30
FT   CHAIN        31   1608       HEMOLYSIN.
SQ   SEQUENCE   1608 AA;  165078 MW;  D669B476FE7DAD51 CRC64;

  Query Match             5.4%;  Score 150.5;  DB 1;  Length 1608;
  Best Local Similarity   20.9%;  Pred. No. 0.37;
  Matches  122;  Conservative   84;  Mismatches  236;  Indels  141;  Gaps   25;

Qy         33 GTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSR--------LAAATSETC 84

Db       1008 GSDLTVDAKGEGGTQRSNSSASQAVTGSIDAAN--GINVNVKKDAIYQGTALNGGRGKTA 1065

Qy         85 LLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA----------------DGTH-------- 120

Db       1066 VNAGGDIRLDQASDKQSESR---SGFNVKASAKGGFTADSKNFGAGFGGGTHNGESSSST 1122

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEA-RT 179

Db       1123 AQVGNISG-QQGVELKAGRDLTLQGTDVKSQGDVSLSAGN--------KVALQAAESTQT 1173

Qy        180 RILAKLDPDNHGGRQPKDVDTRSVGVGSASG-IDDGVVSETHTSTTNSSVRSDPKFWVSV 238

Db       1174 RKESKLSGNIDLGAGSSDSKEKTGGNLSAGGAFDIAKVNESATERQGATIASDGKVTLSA 1233

Qy        239 ---GAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAA---ESATKDQLTQEAFK 292

Db       1234 NGKGDDALHLQGAKVSGGSAAL---------------EAKNGGILLESAKNEQ------- 1271

Qy        293 NPEN------------QKVNIDANGNAIP-----SGELXDDIVEQIAQQAK--------E 327

Db       1272 HKDNWSLGIKANAKGGQTFNKDAGGKVDPNTGKDTHTLGAGLKVGVEQQDKTTHANTGIT 1331

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387

Db       1332 AGDVTLNSGKDTRL-AGARVD------ADSVQGKVGGDLHVESRKDVENGVKVDVDAGLS 1384

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQ------HKVALMPQER-----RRFSDRRDSQGS 436

Db       1385 HSNDPGSSITSKLSKVGTPRYAGKVKEKLEAGVNKVADATTDKYNSVARRLDPQQDTTGA 1444

Qy        437 VASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTG 496

Db       1445 VS---FSKAEGKVTLPATPAG--------EKPQGPLWDRGARTVGGAVKDSITGPAGRQG 1493

Qy        497 RL------IGTPGQGIQSTYALLANSGGLRLGMGGLT--SGGE 531

Db       1494 HLKVNADVVNNNAVGEQSA---IAGKNGVALQVGGQTQLTGGE 1533


RESULT 3
Y456_CHLTR
ID   Y456_CHLTR     STANDARD;      PRT;  1005 AA.
AC   O84462;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein CT456 precursor.
GN   CT456.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/Cx;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0572/CT456/TC0741
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001319; AAC68056.1; -.
DR   PIR; C71513; C71513.
DR   PHCI-2DPAGE; O84462; -.
```

```
KW   Signal; Complete proteome.
FT   SIGNAL        1     40       POTENTIAL.
FT   CHAIN        41   1005       PROTEIN CT456.
SQ   SEQUENCE   1005 AA;  102131 MW;  EC47EC389851CD1E CRC64;

  Query Match             5.4%;  Score 150;  DB 1;  Length 1005;
  Best Local Similarity   22.8%;  Pred. No. 0.21;
  Matches  145;  Conservative   65;  Mismatches  234;  Indels  192;  Gaps   33;

Qy          4 GNLGNNVNGNHLIPPAPPL-PSQTDGAARGGT------GHLISSTGALGSRSLFSPLR-- 54
              ||| | |  |:           |  :  |: | |       || :   |     : |: :
Db        431 GNLPNTVIINNKFKTCVAYGPWNSQEASSGYTPSAWRRGHRVDFGGIFEKANDFNKINWG 490

Qy         55 -----NSMADSVD-SRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPS 108
                  :|  | :  | : ||    |   ||| | |             : |:|
Db        491 TQAGPSSEDDGISFSNETPG--AGP---AAAPSPT----------PSSIPIIN------ 528

Qy        109 AFRVEVQADGTHAAIGEKNGLEVSVT-LSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPM 167
                | |   ||:  ||: |   |: |  :|   |: | ||
Db        529 ---VNVNVGGTNVNIGDTN---VNTTNTTPTTQSTDASTDT------------------ 563

Qy        168 VTVASDIAEARTRILAKLDPDNHGGRQPKDVDTR---SVGVG----------SASGIDDG 214
                |||  :  |              |  |::|    | | |           |:|| | |
Db        564 ----SDIDDINT-----------NNQTDDINTTDKDSDGAGGVNGDISETESSSGDDSG 607

Qy        215 VVSETHTSTTNSSVRSDPKFWVSV--------------------GAIAAG------LAG 247
                || :  |  |:|| :|      :                    | : |       ::
Db        608 SVSSSE-SDKNASVGNDGPAMKDILSAVRKHLDVVYPGENGGSTEGPLPANQTLGDVISD 666

Qy        248 LAATGIAQALALTPE----PDDPTTTDPDQAANAAESAT--------KDQLTQEA----- 290
              :  | ||   |:       ||||||      | ||  |         | ::  |
Db        667 VENKGSAQDTKLSGNTGAGDDDPTTT--AAVGNGAEEITLSDTDSGIGDDVSDTASSSGD 724

Qy        291 ----FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQR 346
                  :| ::     | ||  |||    ||:  : :  :         : |   || |
Db        725 ESGGVSSPSSESNKNTAVGNDGPSGL---DILAAVRKHLDKVYPGDNGGSTEGPLQANQT 781

Qy        347 YED-----QHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT 401
                |      :     :|  :|      | :|:   ||| |:  |  |      |   |||  |
Db        782 LGDIVQDMETTGTSQETVVSP--WKGSTSSTESAGGSGSVQT--LLPSPPPTPSTTTLRT 837

Qy        402 HT---------------VVQQQTGG-IPQHKV--ALMPQERRRFSDRRDSQGSVASTHW 442
                |                ::    || ||       |:|: |       |:|| : ||
Db        838 GTGATTTSLMMGGPIKADIITTGGGGRIPGGGTLEKLLPRIRAHLDISFDAQGDLVSTEE 897

Qy        443 SDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTP 502
                | :|| : :  |:|  |:          :     || :  |  :|:|       |
Db        898 PQLGS-IVNKFRQETGSRGILA-------FVESAPGKPGSA--QVLTGTG-------GDK 940

Qy        503 GQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVN 538
              |   |:  |:   | :  |   |   |:   | ||
Db        941 GNLFQAAAAVTQALGNV-AGKVNLAIQGQKLSSLVN 975


RESULT 4
E74B_DROME
ID   E74B_DROME     STANDARD;      PRT;   883 AA.
AC   P11536; Q9VVI7;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Ecdysone-induced protein 74EF isoform B (ETS-related protein E74B).
GN   EIP74EF OR E74 OR CG6285.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89315191; PubMed=2501755;
RA   Janknecht R., Taube W., Luedecke H.-J., Pongs O.;
RT   "Characterization of a putative transcription factor gene expressed
RT   in the 20-OH-ecdysone inducible puff 74EF in Drosophila
RT   melanogaster.";
RL   Nucleic Acids Res. 17:4455-4464(1989).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=90199900; PubMed=2107982;
RA   Burtis K.C., Thummel C.S., Jones C.W., Karim F.D., Hogness D.S.;
RT   "The Drosophila 74EF early puff contains E74, a complex ecdysone-
RT   inducible gene that encodes two ets-related proteins.";
RL   Cell 61:85-99(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   DEVELOPMENTAL STAGE.
RX   MEDLINE=94038699; PubMed=8223281;
RA   Huet F., Ruiz C., Richards G.;
RT   "Puffs and PCR: the in vivo dynamics of early gene expression during
RT   ecdysone responses in Drosophila.";
RL   Development 118:613-627(1993).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=B; Synonyms=E74B;
CC         IsoId=P11536-1; Sequence=Displayed;
CC       Name=A; Synonyms=E74A;
CC         IsoId=P20105-1; Sequence=External;
CC   -!- DEVELOPMENTAL STAGE: In mid instar larvae salivary glands levels
CC       increase during 86-94 hours of development and represent the
CC       predominant isoform during puff stage 1. Levels remain relatively
CC       constant in late larvae until the premetamorphic pulse of
CC       ecdysone. Transcripts are detected again from puff stages 12-14
CC       and 17-21.
CC   -!- INDUCTION: THE EXPRESSION OF THIS PROTEIN IS DEVELOPMENTALLY
```

```
CC        REGULATED AND IS CORRELATED WITH THE 20-OH-ECDYSONE INDUCED
CC        ACTIVITY OF PUFF 74EF.
CC    -!- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use  by  non-profit  institutions as long  as its content  is  in  no  way
CC    modified and this statement is not removed.  Usage  by  and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; X15087; CAA33195.1; -.
DR    EMBL; M37083; AAA28494.1; -.
DR    EMBL; AE003523; AAF49324.1; -.
DR    PIR; S04722; S04722.
DR    HSSP; P14921; 2STT.
DR    TRANSFAC; T00210; -.
DR    FlyBase; FBgn0000567; Eip74EF.
DR    GO; GO:0006914; P:autophagy; IMP.
DR    InterPro; IPR000418; Ets.
DR    InterPro; IPR002341; HSF_ETS.
DR    Pfam; PF00178; Ets; 1.
DR    PRINTS; PR00454; ETSDOMAIN.
DR    SMART; SM00413; ETS; 1.
DR    PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR    PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR    PROSITE; PS50061; ETS_DOMAIN_3; 1.
KW    Nuclear protein; Transcription regulation; DNA-binding;
KW    Alternative splicing; Developmental protein; Polymorphism.
FT    DOMAIN       78    280       ASP/GLU-RICH (ACIDIC).
FT    DOMAIN      281    759       ALA/GLN/SER-RICH.
FT    DOMAIN      760    883       ARG/LYS-RICH (BASIC).
FT    DNA_BIND    787    869       ETS-DOMAIN.
FT    DOMAIN       27     51       POLY-SER.
FT    DOMAIN       72     79       POLY-ALA.
FT    DOMAIN      289    299       POLY-ALA.
FT    DOMAIN      372    378       POLY-GLN.
FT    DOMAIN      406    415       POLY-GLN.
FT    DOMAIN      434    437       POLY-GLN.
FT    DOMAIN      486    489       POLY-SER.
FT    DOMAIN      510    524       POLY-ALA.
FT    DOMAIN      614    630       POLY-GLN.
FT    DOMAIN      639    642       POLY-GLN.
FT    DOMAIN      651    661       POLY-GLN.
FT    DOMAIN      695    703       POLY-ALA.
FT    DOMAIN      707    711       POLY-SER.
FT    DOMAIN      712    717       POLY-ALA.
FT    DOMAIN      721    733       POLY-ALA.
FT    VARIANT     867    867       H -> Q.
SQ    SEQUENCE   883 AA;  94820 MW;  148D5031A18D1409 CRC64;

  Query Match              5.1%;  Score 144;  DB 1;  Length 883;
  Best Local Similarity   20.8%;  Pred. No. 0.4;
  Matches  124;  Conservative   68;  Mismatches  197;  Indels  208;  Gaps   26;

Qy         39 SSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAATSETCLLGGFEVLHD---- 94
              ||: :| |  | | | ::     | |    |  ||:|   ||| :|      | |: |
Db         38 SSSLSLSSSSSSSLSSATPTPVAS---PVTPTSPPPAAAAPAEASPPAGAELQEDGQQA 94

Qy         95 ---KGPL----DILNTQIGPSAFRVEVQADG----THAAIGEKNGLEVSVTLSPQEWSSL 143
                 : |     |:|           | |:  |    |  : :  ::   |: |  |:   ||
Db         95 KTQEDPTMKDQDMLEKTRQEVKDPVNVEEPGAIVDTESVMARQSPSPVASTKVPE---SL 151

Qy        144 QSIDTEGKNRFVFTGGRGGSGHPMV--------TVASDIAEARTRILAKLDPDNHGGRQP 195
              : |  :                  | |       :|||  | :          :|    |
Db        152 EEISNKS---------------PPVQEDEEESESVASDCREFKVLYNHLRQQQHH--HSP 194

Qy        196 KDVD-TRSVGVGSASGIDD--GVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATG 252
                 | |||          :||    :: |           |||
Db        195 SSPDKTRST-------LDDVSKILWERKQQLQRSSV----------------------- 223

Qy        253 IAQALALTPE---------PDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDA 303
              |  |  | |:              |:  |    | | |      |:  :|:    | ::|  :  :::
Db        224 ITAAPTLQPQQHQQPMSDIEDEETLEDVDDADADVEADAEDEELLEQYQNGYDSPLDLSL 283

Qy        304 NG------------NAIP-------SGELXDDIV--EQIAQQAKEAGEVARQQAVE---- 338
              |             :|:        ||   ||    |:    :   |   :   |   |  :
Db        284 GGATSAAASAAAAASAVSRRRGRTYSGTESDDSAQCERARMRLKPERKAERSAAYKKSLM 343

Qy        339 --------------SNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTT 384
                            | |  ||  :   |  ::|::  |  :|   :                || |
Db        344 KRYYTEIPIVKQSTSPAPQQQLQQQHHLQQQQQQQPHNGSTF------------AGATA 390

Qy        385 ALHRRNQPAEQTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSD 444
               ||  :    ||  |   |    :  |||                     :  ||
Db        391 LLHIK---TEQNTLLTPLQLQQQQ--------------------QQQQG--------- 416

Qy        445 SSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPV---------- 494
                   :: |  ||: | |  :||| :: :        | |     :: |    | :
Db        417 -----LHGAAGNGGSSNGNNAHQQQQPL-----AIPQRPLLHNLLSGGAIHNPHHRNYTT 466

Qy        495 --TGRLIGTPG-QGIQSTYALLANSGG--------LRLGMGGLTSGGETAVSSVNAA 540
                ||     :|    |:     :  ::|||           ||||    ||      | :: ||
Db        467 ATTGSFPPSPADSGVSDVDS--SSSGGQPCADELKARLGMPPATSASAAAAAAAAA 521

RESULT 5
PGCA_MOUSE
ID   PGCA_MOUSE     STANDARD;      PRT;  2132 AA.
AC   Q61282; Q64021;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE   protein) (CSPCP).
GN   AGC1 OR AGC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Cartilage;
RX   MEDLINE=95104847; PubMed=7806222;
RA   Walcz E., Deak F., Erhardt P., Coulter S.N., Fueloep C., Horvath P.,
RA   Doege K.J., Glant T.T.;
RT   "Complete coding sequence, deduced primary structure, chromosomal
RT   localization, and structural analysis of murine aggrecan.";
RL   Genomics 22:364-371(1994).
RN   [2]
RP   SEQUENCE OF 211-326 FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=95004579; PubMed=7920633;
RA   Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A.,
RA   Yamada Y.;
RT   "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in
RT   the aggrecan gene.";
RL   Nat. Genet. 7:154-157(1994).
RN   [3]
RP   INTERACTION WITH FBLN1.
RX   PubMed=10400671;
RA   Aspberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
RT   "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT   versican.";
RL   J. Biol. Chem. 274:20444-20449(1999).
CC   -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC       MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC       IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC       HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC       REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC   -!- SUBUNIT: Interacts with FBLN1.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC       similarity).
CC   -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC       TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
```

```
CC       MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC       CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC       THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC       AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC       AND G3.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC       CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX
CC       DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY
CC       CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD
CC       CAUSES ABSENCE OF AGGRECAN BY TRUNCATION OF THE PROTEIN (MUTATION
CC       IN THE G1 DOMAIN).
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 4 link domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L07049; AAC37670.1; -.
DR   EMBL; S73722; AAB32160.1; -.
DR   EMBL; S73721; AAB32160.1; JOINED.
DR   PIR; A55182; A55182.
DR   HSSP; P98066; 1TSG.
DR   MGD; MGI:99602; Agc1.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR003324; SGXXSG.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF02339; SGXXSG; 60.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 4.
DR   PRINTS; PR01265; LINKMODULE.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   ProDom; PD000918; Link; 4.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00406; IGv; 1.
DR   SMART; SM00445; LINK; 4.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PROSITE; PS01241; LINK; 4.
KW   Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW   Repeat; Immunoglobulin domain.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   2132       AGGRECAN CORE PROTEIN.
FT   DOMAIN       34    147       IG-LIKE V-TYPE.
FT   DOMAIN      170    247       LINK 1.
FT   DOMAIN      268    349       LINK 2.
FT   DOMAIN      504    581       LINK 3.
FT   DOMAIN      602    683       LINK 4.
FT   DOMAIN     1918   2044       C-TYPE LECTIN.
FT   DOMAIN     2048   2106       SUSHI.
FT   DOMAIN       48    140       G1-A.
FT   DOMAIN      152    247       G1-B.
FT   DOMAIN      253    349       G1-B'.
FT   DOMAIN      486    580       G2-B.
FT   DOMAIN      587    682       G2-B'.
FT   DOMAIN      685    803       KS.
FT   DOMAIN      805   1231       CS-1.
FT   DOMAIN     1232   1917       CS-2.
FT   DOMAIN     1917   2132       G3.
FT   DISULFID     51    133       BY SIMILARITY.
FT   DISULFID    175    246       BY SIMILARITY.
FT   DISULFID    199    220       BY SIMILARITY.
FT   DISULFID    273    348       BY SIMILARITY.
FT   DISULFID    297    318       BY SIMILARITY.
FT   DISULFID    509    580       BY SIMILARITY.
FT   DISULFID    533    554       BY SIMILARITY.
FT   DISULFID    607    682       BY SIMILARITY.
FT   DISULFID    631    652       BY SIMILARITY.
FT   DISULFID   1922   1933       BY SIMILARITY.
FT   DISULFID   1950   2042       BY SIMILARITY.
FT   DISULFID   2018   2034       BY SIMILARITY.
FT   DISULFID   2049   2092       BY SIMILARITY.
FT   DISULFID   2078   2105       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    387    387       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    611    611       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    667    667       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1675   1675       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        1171   1173       CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   2132 AA;  222008 MW;  0B2BCDFC6CBDA163 CRC64;

  Query Match             5.0%;  Score 140;  DB 1;  Length 2132;
  Best Local Similarity   21.0%;  Pred. No. 2.2;
  Matches  130;  Conservative   82;  Mismatches  249;  Indels  158;  Gaps   27;

Qy          4 GNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDS 63
              |:|    :|  :   :     :| |   || | | :||  :   |     |  :  |
Db        906 GDLSGLPSGGEITETSTSGAEETSGLPSGGDG-LETSTSGVDDVSGIPTGREGLETSASG 964

Qy         64 -RDIPGLPTNPSRLAAATSE----TCLLGGFEVLHD--KGPLDILNTQIGPSAFRVEVQA 116
               |: |||:       :||     : |  | | |      |  |:     | :  :|  |
Db        965 VEDLSGLPSGEEGSETSTSGIEDISVLPTGGESLETSASGVGDLSGLPSGGES--LETSA 1022

Qy        117 DGTHAAI---GEKNGLEVS------VTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPM 167
               |      |: ||| |      :|: |    ||::  :  ::      || |    :
Db       1023 SGAEDVTQLPTERGGLETSASGVEDITVLPTGRESLETSASGVEDVSGLPSGREG----L 1078

Qy        168 VTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSS 227
               | || | :       : |    |     :|| : | | ||| |       | |: |
Db       1079 ETSASGIED------ISVFPTEAEG-----LDTSASG-GYVSGIPSG---GDGTETSASG 1123

Qy        228 VRSDPKFWVSVGAIAAGLAGL--AATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQ 285
               |       | : :|  ||  :|:|:       |   |:| | : : :    |
Db       1124 VE-------DVSGLPSGGEGLETSASGV--------EDLGPSTRDSLETSASGVDVT--- 1165

Qy        286 LTQEAFK----NPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNA 341
                 |     :||     :  : : :|||                       :: :|::|
Db       1166 ----GFPSGRGDPETSVSGVGDDFSGLPSG---------------------KEGLETSA 1199

Qy        342 QAQQRYEDQHARRQEELQLSSG----------------------------------IGY 366
                :     : ::: :  :||                                    ||
Db       1200 SGAEDLSGLPSGKEDLVGSASGALDFGKLPPGTLGSGQTPEVNGFPSGFSGEYSGADIGS 1259

Qy        367 GLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIP-QHKVALMPQERR 425
               | || |   |: :|  |     :   |  | ||:  :  :  | ||    :|: :|
Db       1260 GPSSGLPDFSGLPSGFPTVSLVDSTLVEVITATTSSELEGRGTIGISGSGEVSGLP--LG 1317

Qy        426 RFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP--GYS 483
                  |  | :  |  |  :|   :   |  |               : :|:  |
Db       1318 ELDSSADISGLPSGTELSGQASGSPDSSGETSG-------------FFDVSGQPFGSSG 1363

Qy        484 VIQNFSGSGPVTGRLIGTPGQGIQSTYALL----------ANSGGLRLGMG---GLTS- 528
               | :  ||   ::|:  |||      |    |             :  |:  | |   |:||
Db       1364 VSEETSGIPEISGQPSGTPDTTATSGVTELNELSSGQPDVSGDGSGILFGSGQSSGITSV 1423
```

```
Qy        529 GGETA-VSSVNAAPTPGPV 546
              |||: :| ::  |:  ||
Db       1424 SGETSGISDLSGQPSGFPV 1442

RESULT 6
IGA_NEIGO
ID   IGA_NEIGO      STANDARD;      PRT;  1532 AA.
AC   P09790;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   IgA-specific serine endopeptidase precursor (EC 3.4.21.72) (IgA
DE   protease).
GN   IGA.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=MS11;
RX   MEDLINE=87115823; PubMed=3027577;
RA   Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT   "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT   IgA protease.";
RL   Nature 325:458-462(1987).
RN   [2]
RP   ACTIVE SITE.
RX   MEDLINE=90154052; PubMed=2105953;
RA   Bachovchin W.W., Plaut A.G., Flentke G.R., Lynch M., Kettner C.A.;
RT   "Inhibition of IgA1 proteinases from Neisseria gonorrhoeae and
RT   Hemophilus influenzae by peptide prolyl boronic acids.";
RL   J. Biol. Chem. 265:3738-3743(1990).
CC   -!- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC   -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC       certain Pro-|-Xaa bonds in the hinge region. No small molecule
CC       substrates are known.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC       SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC       OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC       DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X04835; CAA28538.1; -.
DR   PIR; A26039; A26039.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   InterPro; IPR000710; IgA_S6.
DR   InterPro; IPR004899; Pertactin.
DR   Pfam; PF03797; Autotransporter; 1.
DR   Pfam; PF02395; IGA1; 1.
DR   Pfam; PF03212; Pertactin; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW   Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;
KW   Transmembrane; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    986       IGA-SPECIFIC SERINE ENDOPEPTIDASE.
FT   PROPEP      987   1532       HELPER PEPTIDE.
FT   ACT_SITE    278    278       POTENTIAL.
FT   SITE        986    987       CLEAVAGE (AUTO-).
FT   SITE       1018   1019       CLEAVAGE (AUTO-).
FT   SITE       1121   1122       CLEAVAGE (AUTO-).
SQ   SEQUENCE   1532 AA;  168976 MW;  68FF4112BD22F40D CRC64;

  Query Match              4.9%;  Score 137;  DB 1;  Length 1532;
  Best Local Similarity   21.0%;  Pred. No. 2.1;
  Matches  102;  Conservative   60;  Mismatches  183;  Indels  140;  Gaps   20;

Qy         34 TGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLH 93
              ||::  :|| |  ::|                :| |:  ||          |  :  | |  :
Db        782 TGYVTCNTGNLSDKAL---------NSFDATRINGNVNLNQNAALVLGKAALWGKIQ--- 829

Qy         94 DKGPLDILNTQIGPSAFRVEVQADGTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNR 153
                          |  || :            |:    :|: |    ::     :  |
Db        830 ------------GQGNSRVSLNQHSKWHLTGDSQVHNLSLADSHIHLNNASDAQSANKYH 877

Qy        154 FVFTGGRGGSG--HPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI 211
                 :     |:|  | :  :| :: :   ::| |    ::   |       |  :: |  :  |:
Db        878 TIKINHLSGNGHFHYLTDLAKNLGD---KVLVK---ESASGHYQLHVQNKT-GEPNQEGL 930

Qy        212 D--DGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAAT-----GIAQALALTPEPD 264
              |  |             |||:   : :||:          |   |   |      || :
Db        931 DLFDA-----------SSVQDRSRLFVSLANHYVDLGALRYTIKTENGITRLYNPYAGNG 979

Qy        265 DPTTTDPDQAANAAESATKDQLTQEA-FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQ 323
              |     |  |||  |  |   |   |   |     |:|  |           :   |   |
Db        980 RPVKPAPSPAANTASQAQKATQTDGAQIAKPQNIVV--------APPSPQANQAEEALRQ 1031

Qy        324 QAKEAGEVARQQAVESNAQAQQRYED------QHARRQEELQLSSGIGYGLSSALIVAGG 377
              ||| | :| |||| |:   |:|: |:        : ||:|||  :            :| :  |
Db       1032 QAK-AEQVKRQQAAEAEKVARQKDEEAKRKAAEIARQQEEAR---------KAAELAAKQ 1081

Qy        378 IGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRR----------- 426
                       |     | ||: :                   |:    |: |||
Db       1082 KAEAERKARELARQKAEEAS-----------------HQANAKPKRRRRRAILPRPPAPV 1124

Qy        427 FS----DRRDSQGSVASTHWSDSS-----------SEVVNPYAEVGGARNSLSAHQPEE 470
              ||      |  :|:                 |:||           |::| | |:             ||
Db       1125 FSLDDYDAKDN---------SESSIGNLARVIPRMGRELINDYEEI-----------PLE 1164

Qy        471 HIYDE 475
                 : ||
Db       1165 ELEDE 1169

RESULT 7
YPJA_ECOLI
ID   YPJA_ECOLI     STANDARD;      PRT;  1569 AA.
AC   P52143; P76610; P77017; P77019;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical outer membrane protein ypjA.
GN   YPJA OR B2647.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
```

```
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U36840; AAA79815.1; ALT_SEQ.
DR   EMBL; AE000350; AAC75695.1; -.
DR   EMBL; D90889; BAA16514.1; ALT_INIT.
DR   EMBL; D90890; BAA16518.1; ALT_INIT.
DR   PIR; A65044; A65044.
DR   EcoGene; EG13213; ypjA.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   InterPro; IPR004899; Pertactin.
DR   InterPro; IPR003991; Pertactin_C.
DR   Pfam; PF03797; Autotransporter; 1.
DR   Pfam; PF03212; Pertactin; 1.
DR   PRINTS; PR01484; PRTACTNFAMLY.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW   Hypothetical protein; Outer membrane; Complete proteome.
SQ   SEQUENCE   1569 AA;  162774 MW;  0D006A9A32154596 CRC64;

  Query Match             4.9%;  Score 136;  DB 1;  Length 1569;
  Best Local Similarity   20.9%;  Pred. No. 2.5;
  Matches  127;  Conservative   69;  Mismatches  259;  Indels  154;  Gaps   26;

Qy         19 APPLPSQTDGAARGGT--GHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRL 76
              |  :  |: ||    |  | ||  | : |       :||| | :|       :  |   |
Db        579 AKHVEQQSGGALIASTTSGTLIEGTNSYGDAFY---IRNSEAKNV-------VLENAGSL 628

Qy         77 AAATSETCLLGGFEVLHDKGPLDILNTQIGP---SAFRVEVQADGTHAAIGEKNGLEVSV 133
                 |    :    :::  | :|:    :|    ||    : |  |       | | :
Db        629 TVVTGSRAV---DTIINANGKMDVYGKDVGTVLNSAGTQTIYASATSDKANIKGGKQTVY 685

Qy        134 TLSPQE--WSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTRILAKLDPD--- 188
              |: :    |  | :|     :    ||     :    : :||     :|:|    :
Db        686 GLATEANIESGEQIVDGGSTEKTHINGGTQTVQNYGKAINTDIVSGLQQIMANGTAEGSI 745

Qy        189 --------NHGGRQPKDV-------DTRSVGVGSASGI---DDGVVSETHTSTTNSSVRS 230
                    | ||      |       | |    |||:||     | :  |  :|  :  |:
Db        746 INGGSQVVNEGGLAENSVLNDGGTLDVRE--KGSATGIQQSSQGALVATTRATRVTGTRA 803

Qy        231 DPKFWVSVGAIAAGLAGLAATGI--AQALALTPEPDDPTTTDPDQAANAAESATKDQLTQ 288
              |         :|  :   ||  |  |    || | |  |::|  |          | : |
Db        804 D--------GVAFSIEQGAANNILLANGGVLTVESD--TSSDKTQVNMGGREIVKTKAT- 852

Qy        289 EAFKNPENQKVNIDANGNAIPSGELXDDIVEQIA----------QQAKEAGEVARQQ--- 335
                        | |  :  ||    |||  :|           |      ||  : :
Db        853 -------------ATGTTLTGGE---QIVEGVANETTINDGGIQTVSANGEAIKTKINE 895

Qy        336 ----AVESNAQAQQRYEDQHARRQEE----LQLSSGIGYGLSS----------------A 371
                  |  | :|    ::  |  |      :::|    ||  |
Db        896 GGTLTVNDNGKATDIVQNSGAALQTSTANGIEISGTHQYGTFSISGNLATNMLLENGGNL 955

Qy        372 LIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRR 431
              |::||       |       |   | : |  :      :||  |: :      | :  :
Db        956 LVLAGTEARDSTVGKGGAMQNLGQDSATKVN------SGG--QYTLGRSKDEFQALARAE 1007

Qy        432 DSQ--GSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFS 489
                  | | |  |  :    :       | | ||  |||   | :::
Db       1008 DLQVAGGTAIVYAGTLAD------ASVSGATGSLSLMTPRDNV----------TPVKLE 1050

Qy        490 GSGPVTGRLIGTPGQGIQSTYALL--ANSGGLRLGMGGLTSG--------------GET 532
                |:  :|     | | |: :| | |  |: | :  |       :|            |:
Db       1051 GAVRITDSATLTLGNGVDTTLADLTAASRGSVWLNSNNSCAGTSNCEYRVNSLLLNDGDV 1110

Qy        533 AVSSVNAAP 541
                :|:  |||
Db       1111 YLSAQTAAP 1119


RESULT 8
PGCA_RAT
ID   PGCA_RAT       STANDARD;      PRT;  2124 AA.
AC   P07897;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE   protein) (CSPCP).
GN   AGC1 OR AGC.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88087070; PubMed=3693370;
RA   Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RT   "Complete primary structure of the rat cartilage proteoglycan core
RT   protein deduced from cDNA clones.";
RL   J. Biol. Chem. 262:17757-17767(1987).
RN   [2]
RP   REVISION TO 698.
RA   Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL   J. Biol. Chem. 263:10040-10040(1988).
RN   [3]
RP   SEQUENCE OF 1856-2124 FROM N.A.
RX   MEDLINE=86250698; PubMed=2424893;
RA   Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT   "Partial cDNA sequence encoding a globular domain at the C terminus
RT   of the rat cartilage proteoglycan.";
RL   J. Biol. Chem. 261:8108-8111(1986).
CC   -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC       MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC       IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC       HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC       REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC   -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC       similarity).
CC   -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC       TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC       MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC       CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC       THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC       AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC       AND G3.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC       CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 4 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M13518; AAA41836.1; -.
DR   EMBL; J03485; AAA21000.1; ALT_SEQ.
DR   PIR; A92623; A28452.
DR   HSSP; P98066; 1TSG.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR003324; SGXXSG.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF02339; SGXXSG; 55.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 4.
DR   PRINTS; PR01265; LINKMODULE.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   ProDom; PD000918; Link; 4.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00406; IGv; 1.
DR   SMART; SM00445; LINK; 4.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PROSITE; PS01241; LINK; 4.
KW   Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW   Repeat; Immunoglobulin domain.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   2124       AGGRECAN CORE PROTEIN.
FT   DOMAIN       34    147       IG-LIKE V-TYPE.
FT   DOMAIN      170    247       LINK 1.
FT   DOMAIN      268    349       LINK 2.
FT   DOMAIN      504    581       LINK 3.
FT   DOMAIN      602    683       LINK 4.
FT   DOMAIN     1910   2036       C-TYPE LECTIN.
FT   DOMAIN     2040   2098       SUSHI.
FT   DOMAIN       48    140       G1-A.
FT   DOMAIN      152    247       G1-B.
FT   DOMAIN      253    349       G1-B'.
FT   DOMAIN      486    580       G2-B.
FT   DOMAIN      587    682       G2-B'.
FT   DOMAIN      685    798       KS.
FT   DOMAIN      801   1226       CS-1.
FT   DOMAIN     1227   1909       CS-2.
FT   DOMAIN     1910   2124       G3.
FT   DISULFID     51    133       BY SIMILARITY.
FT   DISULFID    175    246       BY SIMILARITY.
FT   DISULFID    199    220       BY SIMILARITY.
FT   DISULFID    273    348       BY SIMILARITY.
FT   DISULFID    297    318       BY SIMILARITY.
FT   DISULFID    509    580       BY SIMILARITY.
FT   DISULFID    533    554       BY SIMILARITY.
FT   DISULFID    607    682       BY SIMILARITY.
FT   DISULFID    631    652       BY SIMILARITY.
FT   DISULFID   1914   1925       BY SIMILARITY.
FT   DISULFID   1942   2034       BY SIMILARITY.
FT   DISULFID   2010   2026       BY SIMILARITY.
FT   DISULFID   2041   2084       BY SIMILARITY.
FT   DISULFID   2070   2097       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    387    387       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    611    611       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    667    667       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1842   1842       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2124 AA;  221117 MW;  E30BBE61593A34B1 CRC64;
```

```
  Query Match             4.8%;  Score 135.5;  DB 1;  Length 2124;
  Best Local Similarity   20.7%;  Pred. No. 3.9;
  Matches  138;  Conservative   77;  Mismatches  204;  Indels  247;  Gaps   33;

Qy         17 PPAPPLPSQTD----GAARGGTGHL--------ISSTGALGSRSLFSPLRNSMADSVDSR 64
              |    |||  :      |:  ||| |        |: | | |                 :
Db        880 PTVDRLPSGGESLEGSASASGTGDLSGLPSGGEITETSASG----------------TE 922

Qy         65 DIPGLPTNPSRLAAATSETCLLGGFEVL-HDKGPLDILNTQIGPSAFRVEVQADGTHAAI 123
              :| |||:    |  :||    : |  ||   :| |               |  | |
Db        923 EISGLPSGGDDLETSTSG---IDGASVLPTGRGGL-------------ETSASGVEDLS 965

Qy        124 GEKNGLEVSVTLSPQEWSSLQSIDT--EGKNRFVFTGGRG------GSGHPMVTVASDIA 175
              |  :| | | | :    | :: |      |::       | |         |  : | || :
Db        966 GLPSGEEGSETST----SGIEDISVLPTGESPETSASGVGDLSGLPSGGESLETSASGVE 1021

Qy        176 EARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSET---HTSTTNSSVRSDP 232
              :  |::     |   || :                |||||:|   |  |   :  |: | |
Db       1022 DV-TQL-----PTERGGLET-----------SASGIEDITVLPTGRENLETSASGVE--- 1061

Qy        233 KFWVSVGAIAAGLAGL--AATGI---------AQALALT---------PEPDDPTTTDP- 271
                |  : :|  ||  :|:||          |: |  :           |  :| | |
Db       1062 ----DVSGLPSGKEGLETSASGIEDISVFPTEAEGLETSASGGYVSGIPSGEDGTETSTS 1117

Qy        272 --------------DQAANAAES---ATKDQLTQEAFKNPENQKVNIDANGNAIPSGE- 312
                           : :|:  |      |:| |         |     :|  |   |||
Db       1118 GVEGVSGLPSGGEGLETSASGVEDLGLPTRDSL--------ETSASGVDVTG--YPSGRE 1167

Qy        313 --------LXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSG- 363
                     : ||:            :|  : |: :|::|    :       : ::: :  :||
Db       1168 DTETSVPGVGDDL----------SGLPSGQEGLETSASGAEDLGGLPSGKEDLVGSASGA 1217

Qy        364 ---------------------------------IGYGLSSALIVAGGIGAGVTTALHRR 389
                                               || | || |    |: :| |
Db       1218 LDFGKLPSGTLGSGQTPEASGLPSGFSGEYSGVDIGSGPSSGLPDFSGLPSGFPTVSLVD 1277

Qy        390 NQPAEQTTTTTTHTVVQQQT---------GGIPQHKVALMPQERRRFSDRRDSQGSVAST 440
              :   |  | ||   :  : |          | |  ::           |  |  : |
Db       1278 STLVEVITATTASELEGRGTISVSGSGEESGPPLSEL----------DSSADISGLPSGT 1327

Qy        441 HWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIG 500
                |  :|  ::   |  |                : :|:  |  |    ||:|  |  :
Db       1328 ELSGQTSGSLDVSGETSG-------------FFDVSGQPFGS-----SGTGEGTSGIPE 1368

Qy        501 TPGQGIQS--TYALLANSG-------------GLRLGMG---GLTS-GGETA-VSSVNAA 540
                || ::|  |  :   ||             |:  | |   |:|| |||: :| ::
Db       1369 VSGQAVRSPDTTEISELSGLSSGQPDVSGEGSGILFGSGQSSGITSVSGETSGISDLSGQ 1428

Qy        541 PTPGPV 546
                |:  ||
Db       1429 PSGFPV 1434
```

```
RESULT 9
APMU_PIG
ID   APMU_PIG       STANDARD;      PRT;  1150 AA.
AC   P12021;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Apomucin (Mucin core protein) (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=91236743; PubMed=2033060;
RA   Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT   "Porcine submaxillary mucin contains a cystine-rich,
RT   carboxyl-terminal domain in addition to a highly repetitive,
```

```
RT   glycosylated domain.";
RL   J. Biol. Chem. 266:9678-9686(1991).
RN   [2]
RP   SEQUENCE OF 1-503 FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=88087170; PubMed=2826455;
RA   Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT   "Porcine submaxillary gland apomucin contains tandemly repeated,
RT   identical sequences of 81 residues.";
RL   J. Biol. Chem. 263:1081-1088(1988).
RN   [3]
RP   SEQUENCE OF 45-80.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=87280230; PubMed=3611111;
RA   Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A.,
RA   Johnson W.C. Jr., Hill R.L.;
RT   "Structural properties of porcine submaxillary gland apomucin.";
RL   J. Biol. Chem. 262:11339-11344(1987).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=97248516; PubMed=9092502;
RA   Gerken T.A., Owens C.L., Pasumarthy M.;
RT   "Determination of the site-specific O-glycosylation pattern of the
RT   porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
RT   for the polypeptide:galnac transferase peptide binding site.";
RL   J. Biol. Chem. 272:9709-9719(1997).
CC   -!- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
CC       SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
CC       GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
CC       THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
CC       ENVIRONMENT.
CC   -!- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
CC       MULTIMERIC MUCIN STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC   -!- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
CC       RESIDUES.
CC   -!- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
CC       THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
CC       RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
CC       GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE
CC       PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
CC       ENHANCE GLYCOSYLATION.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
CC   -!- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61883; AAA30998.1; -.
DR   EMBL; M21174; AAA30990.1; -.
DR   InterPro; IPR006208; Cys_knot.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00007; Cys_knot; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Glycoprotein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1    368       81 AA TANDEM REPEATS.
FT   REPEAT       <1     44       1.
FT   REPEAT       45    125       2.
FT   REPEAT      126    206       3.
FT   REPEAT      207    287       4.
FT   REPEAT      288    368       5.
FT   REPEAT      369    391       6 (INCOMPLETE).
FT   DOMAIN      929    995       VWFC.
FT   DOMAIN     1062   1146       CTCK.
FT   DISULFID   1062   1109       BY SIMILARITY.
FT   DISULFID   1076   1123       BY SIMILARITY.
FT   DISULFID   1085   1139       BY SIMILARITY.
FT   DISULFID   1089   1141       BY SIMILARITY.
FT   DISULFID      ?   1145       BY SIMILARITY.
FT   CARBOHYD     46     46       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     50     50       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     51     51       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     57     57       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     58     58       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     61     61       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     66     66       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     67     67       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     73     73       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     74     74       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     76     76       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     77     77       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     81     81       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     83     83       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     87     87       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     91     91       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     93     93       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     94     94       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     96     96       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     98     98       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    101    101       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    103    103       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    104    104       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    106    106       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    107    107       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    108    108       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    110    110       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    114    114       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    117    117       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    123    123       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    124    124       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    917    917       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    985    985       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1002   1002       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1150 AA;  109615 MW;  3CB68B5D29DD7F5A CRC64;
```

Query Match 4.8%; Score 134.5; DB 1; Length 1150;
Best Local Similarity 22.1%; Pred. No. 2;
Matches 126; Conservative 60; Mismatches 260; Indels 125; Gaps 26;

```
Qy         24 SQTDGAARGGTGHLISSTGALGS--RSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATS 81
Db         91 SGTTGTVSGASGSTGSSSGSPGATGASIGQP-ETSRISVAGSSGAPAVSSGASQ-AAGTS 148

Qy         82 ETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTHAAIGEKNGLEVSVTLSPQEWS 141
Db        149 -----------GAGPGTTASSVGVTETARPSVAGSGTTGTVSGASGSTGSSSGSPGATG 196

Qy        142 SLQSIDTEGKNRFVFTGGRGG------------SGHPMVTVASDIAEARTRILAKLDPD 188
Db        197 A--SIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTET-----ARPS 249

Qy        189 NHGGRQPKDVDTRSVGVGSASGIDDGV---VSETHTSTTNSSVRSDPKFWVSVGAIAAGL 245
Db        250 VAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAAGT 309

Qy        246 AG------LAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKV 299
Db        310 SGAGPGTTASSVGVTETARPSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRI 369

Qy        300 NI--DANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEE 357
```

```
                ::   :   |: ||          | ||  ||      ::|
Db        370 SVAGSSGAPAVSSG----------ASQA--AGTSEATTSIEGAG---------------- 401

Qy        358 LQLSSGIGYGLSSALIVAG-------GIGAGVTTALHRRN-QP----AEQTTTT----TT 401
                :||:|:  : |    |         ||  | |  : |:  :|     | ||:     ||
Db        402 ---TSGVGF-KTEATTFPGENETTRVGIATGTTGIVSRKTLEPGSYNTEATTSIGRSGTT 457

Qy        402 HTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARN 461
              ||        :|     ::|       | :    ||  :|  |  |         |
Db        458 HT-------DLPGGTTIVLPGFSH--SSQSSKPGSSVTTPGSPESGSETGTSGEFSTTVI 508

Qy        462 SLSAHQPEEHIYDEVAADPGYSVIQNFSGSGP-VTGRLIGTP-GQGIQSTYALLANSGGL 519
              | |:|  |   :   :  ||       :|| |  || | ||    |  :| |    |
Db        509 SGSSH-TEATTFIGGSGSPG-------TGSRPGTTGELSGTTIASGNATTEA--TTSTET 558

Qy        520 RLG---------MGGLTSGGETAVSSVNAAP 541
              |:|         |   || ||  |   : |
Db        559 RIGPQTGAQTTVPGSQVSGSETGTSEAVSNP 589


RESULT 10
SLAP_CAUCR
ID   SLAP_CAUCR     STANDARD;      PRT;  1025 AA.
AC   P35828; Q46015; Q9RFI2;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   S-layer protein (Paracrystalline surface layer protein).
GN   RSAA OR CC1007.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=93007489; PubMed=1393820;
RA   Gilchrist A., Fisher J.A., Smit J.K.;
RT   "Nucleotide sequence analysis of the gene encoding the Caulobacter
RT   crescentus paracrystalline surface layer protein.";
RL   Can. J. Microbiol. 38:193-202(1992).
RN   [2]
RP   REVISIONS TO 376; 636 AND 842-843.
RA   Awram P.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JS3001;
RA   Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;
RT   "The secretion signal of C. crescentus S-layer protein is located in
RT   the C-terminal 82 amino acids of the molecule.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN   [5]
RP   SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=89008089; PubMed=3049545;
RA   Fisher J.A., Smit J.K., Agabian N.;
RT   "Transcriptional analysis of the major surface array gene of
RT   Caulobacter crescentus.";
RL   J. Bacteriol. 170:4706-4713(1988).
RN   [6]
RP   CHARACTERIZATION.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=98292737; PubMed=9620954;
RA   Awram P., Smit J.K.;
RT   "The Caulobacter crescentus paracrystalline S-layer protein is
RT   secreted by an ABC transporter (type I) secretion apparatus.";
RL   J. Bacteriol. 180:3062-3069(1998).
CC   -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC       OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
CC       PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
CC       (TYPE I) SECRETION APPARATUS.
CC   -!- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC       SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC       SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF062345; AAC38665.2; -.
DR   EMBL; AF193063; AAF19365.1; -.
DR   EMBL; AE005779; AAK22991.1; ALT_INIT.
DR   PIR; A48995; A48995.
DR   HSSP; P22629; 1SWC.
DR   TIGR; CC1007; -.
DR   InterPro; IPR001343; Hemlysn_Ca_bind.
DR   Pfam; PF00353; hemolysinCabind; 3.
DR   PRINTS; PR00313; CABNDNGRPT.
KW   Cell wall; S-layer; Calcium-binding; Complete proteome.
FT   INIT_MET      0      0
SQ   SEQUENCE   1025 AA;  98001 MW;  AD7A326E1363D8AC CRC64;

  Query Match             4.7%;  Score 132.5;  DB 1;  Length 1025;
  Best Local Similarity   21.0%;  Pred. No. 2.3;
  Matches  117;  Conservative   66;  Mismatches  232;  Indels  143;  Gaps   24;

Qy         26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCL 85
              || ||         |:|  ||         |:    |:  : |  | : :|   :
Db        229 TDNAAGVNLFTAYPSSGVSGSTL-------SLTTGTDT--LTGTANNDTFVAGEVAGAAT 279

Qy         86 LGGFEVLHDKGPLDILN----------TQIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134
              |   : |       |:||          | :  |          |    :   :|:
Db        280 LTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSGVTGLTA 339

Qy        135 LSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTRILAKLDPDNHGGRQ 194
              |:     : |::          |     : :  |   :::  | | : :
Db        340 LNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVASTGVTS---------- 388

Qy        195 PKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIA 254
                 | :||  ||:    | ||    |  |||         : ||||  : |  |  :|
Db        389 ----GTTTVGANSAA---SGTVS---VSVANSST-------TTTGAIA--VTGGTAVTVA 429

Qy        255 QALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIPSGELX 314
              |          : | |  |        | |   :|| |             |   |  :| :
Db        430 QTAG---NAVNTTLTQADVTVTGNSSTTAVTVTQTA----------AATAGATVAGRVN 475

Qy        315 D--DIVEQIAQQAKEAGEVAR-------QQAVESNAQAQQRYEDQHARRQEELQLSSGIG 365
                 | :  |  |  ||::|            ::|:|                  || |
Db        476 GAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALTTVNL----------SGTG 521

Qy        366 YGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERR 425
                |        |||  |  ||      |   | |   : :    || |   : |
Db        522 TSL--------GIGRGALTA-----TPTANTLTLNVNGLT--TTGAITDSEAA------- 559

Qy        426 RFSDRRDSQGSVASTHWSDSSSEVVNPYAEV----GGARNSLSAHQPEEHIYDEVAADPG 481
                :|   :  ::| :  | : : :|   |       | || ::::|           ||  |
```

```
Db        560 --ADDGFTTINIAGSTASSTIASLVAADATTLNISGDARVTITSH--------TAAALTG 609

Qy        482 YSVIQNFS---GSGPVTGRLIGTPGQGIQS------TYALLANSG-------GLRLGMGG 525
Db        610 ITVTNSVGATLGAELATG-LVFTGGAGADSILLGATTKAIVMGAGDDTVTVSSATLGAGG 668

Qy        526 LTSGG---ETAVSSVNAA 540
Db        669 SVNGGDGTDVLVANVNGS 686


RESULT 11
FLID_VIBCH
ID   FLID_VIBCH     STANDARD;      PRT;   666 AA.
AC   Q9KQ63;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Flagellar hook-associated protein 2 (HAP2) (Filament cap protein)
DE   (Flagellar cap protein).
GN   FLID OR VC2140.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF
CC       THE FLAGELLAR FILAMENT BY FACILITATING POLYMERIZATION OF THE
CC       FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CAPPING
CC       STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH
CC       THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT
CC       POLYMERIZATION AT THE DISTAL END (BY SIMILARITY).
CC   -!- SUBUNIT: Homopentamer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Flagellar.
CC   -!- SIMILARITY: BELONGS TO THE FLID FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004287; AAF95285.1; -.
DR   PIR; G82111; G82111.
DR   TIGR; VC2140; -.
DR   InterPro; IPR003481; FliD.
DR   Pfam; PF02465; FliD; 1.
KW   Flagella; Coiled coil; Complete proteome.
FT   DOMAIN      354    419       COILED COIL (POTENTIAL).
SQ   SEQUENCE   666 AA;  72330 MW;  6794EFEC34A2A0D2 CRC64;

  Query Match             4.7%;  Score 132;  DB 1;  Length 666;
  Best Local Similarity   22.4%;  Pred. No. 1.4;
  Matches  148;  Conservative   91;  Mismatches  249;  Indels  174;  Gaps   37;

Qy          2 PIG-NLGNNVNGN-HLIPPAPPLPSQTD-GAARGGTGHLISSTGALGS-----RSLFSPL 53
Db          5 PMGMNTGFDINGMVSKIVSAERVPKQQRIDNERTNIDTSISAYGRLRESLDTMKNLMTQF 64

Qy         54 RNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGG---------------FEVLHDK-- 95
Db         65 RQEKAFAVRKVD-----TSNEQVVSATATTEAIAGNYSVDVLQLAQSHKIASEVL-DKDA 118

Qy         96 --GPLDILNTQIGPSAFRVEVQADGTHAAI-----GEKN--GLEVSVTLSPQEWSSLQSI 146
Db        119 KFGP-GKLHISLGDKSFTLDVQGNSKLVDIVRGINGEKSNPGVRASIINDVEGPRLIVAS 177

Qy        147 DTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR----------ILAKLDPDNHGGRQP 195
Db        178 NVSGKDHSVKMSAQAEPGNPLKQLEYKTLEQRVRDLEKARAQAQQLIAPLTPE-----QQ 232

Qy        196 KDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQ 255
Db        233 KVAAKVAEKIGDAARLVDQEVAQEIRSAAQS--------------AQGAAGEALNA--- 274

Qy        256 ALALTPEPDDPTTTDPDQAANAAESATK-----DQL----------TQEAFKNPENQKVN 300
Db        275 --------GELTESAVKAAANAASEAKKYIRPEDRIPGWTETASGTLLDSYWEPEEE--- 323

Qy        301 IDANGNA----IP------SGELXDD-IVEQIAQQAKE---AGEVARQQAV--------- 337
Db        324 LDAQGQKKAADVPGWSNTASGTLLDSYVTPQEAQQKLEQKLAQEKAQIEAAIRSGKMTPE 383

Qy        338 ESNAQA-------QQRYEDQHARRQEELQLSSGI--GYG-------LSSALIVAGGIGAG 381
Db        384 EAKAQARAKLSPEERAYIEQVEKAQAALNAAQSAFDGYGGMTEVQSAQDSMVVLDGV--- 440

Qy        382 VTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVAS-T 440
Db        441 --ATLSSNNNIIENAIEGVNLTLKGKTDRNQPPAEIGI---EYDRERVRNDIEQFVAAYN 495

Qy        441 HWSDSSSEV--VNPYAEVGG--ARNSL--SAHQPEEHIYDEVAADPGYSVIQNFSGSGPV 494
Db        496 QFFQTSKELAGVDPRTGQAGPLAGDSIVRSADSRLKTVFSS-------SIEQAPENLKSL 548

Qy        495 TGRLIGTPGQG-IQSTYALL---ANSGGLRL-----GMGGLTSGGETAVSSVNAAPTPGP 545
Db        549 TEFGITTTRQGTLEINYAMLDRQLNNNFTKLGEFFGGNQGFAKRVEDAISSMTG--VTGS 606

Qy        546 VR 547
Db        607 IR 608


RESULT 12
CSG_HALVO
ID   CSG_HALVO      STANDARD;      PRT;   827 AA.
AC   P25062;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cell surface glycoprotein precursor (S-layer protein).
OS   Halobacterium volcanii (Haloferax volcanii).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Haloferax.
OX   NCBI_TaxID=2246;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91072265; PubMed=2123862;
RA   Sumper M., Berg E., Mengele R., Strobel I.;
RT   "Primary structure and glycosylation of the S-layer protein of
RT   Haloferax volcanii.";
RL   J. Bacteriol. 172:7111-7118(1990).
CC   -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC       OF PROTEINS WHICH COAT THE SURFACE OF THE CELL.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This archaea is covered by an S-
CC       layer with hexagonal symmetry.
CC   -!- PTM: PROTEIN IS CONTAINS O-LINKED GLYCANS WHICH CONSIST OF GLC-GAL
CC       DISACCHARIDES.
CC   -!- SIMILARITY: TO H.HALOBIUM AND H.JAPONICA CELL SURFACE
CC       GLYCOPROTEINS.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M62816; AAA72996.1; -.
DR   PIR; A37849; A37849.
DR   GlycoSuiteDB; P25062; -.
KW   Glycoprotein; Transmembrane; Cell wall; S-layer; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35    827       CELL SURFACE GLYCOPROTEIN.
FT   DOMAIN      760    803       THR-RICH.
FT   TRANSMEM    804    823       POTENTIAL.
FT   CARBOHYD     47     47       N-LINKED (GLC. . .) (POTENTIAL).
FT   CARBOHYD    117    117       N-LINKED (GLC. . .) (POTENTIAL).
FT   CARBOHYD    308    308       N-LINKED (GLC. . .) (POTENTIAL).
FT   CARBOHYD    313    313       N-LINKED (GLC. . .) (POTENTIAL).
FT   CARBOHYD    404    404       N-LINKED (GLC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLC. . .) (POTENTIAL).
FT   CARBOHYD    766    766       N-LINKED (GLC. . .) (POTENTIAL).
SQ   SEQUENCE   827 AA;  85188 MW;  B208E96283699974 CRC64;

  Query Match             4.7%;  Score 132;  DB 1;  Length 827;
  Best Local Similarity   21.4%;  Pred. No. 1.9;
  Matches  124;  Conservative   81;  Mismatches  243;  Indels  132;  Gaps   26;

Qy          2 PIGNLGNNVNGNHLIPPAPPL-PSQTDG---AARGG--TGHLISSTGALGSRSLFSPLRN 55
              |:|    : :|:   |    | | ||     : ||  ||  |||: | |:  : :
Db        101 PLGTYDVDGSGSATTPNVTLLAPRITDSEILTSSGGDVTGSAISSSDA-GNLYVNADYNY 159

Qy         56 SMADSVD-------SRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPS 108
                |: |:         ||             |:|  :| | :   | | :   ::  |
Db        160 ESAEKVEVTVEDPSGTDI------------TNE--VLSGTDTFVDDGSIGSTSSTGGGV 204

Qy        109 AFRVEVQADGTHAAIGEKNGLE----------VSVTLSPQEWSSLQSIDTEGKN-----R 153
               :  | | :  | |  | |          :::|:| |:   :: :|:|          :
Db        205 GIDMSDQDAGEYTIILE--GAEDLDFGDATETMTLTISSQDEIGIE-LDSESVTQGTDVQ 261

Qy        154 FVFTGGRGGSGHPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDD 213
              :  | |  |: | :    ||:     |              | |:|  |::|  |  || :
Db        262 YTVTNGIDGNEHVVAMDLSDLQNDATT-----------EQAKEV-FRNIGDTSEVGIAN 308

Qy        214 GVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQ 273
               :  |:|  :| :        :      : | :| |  :   |      |     |
Db        309 SSATNTSGSSTGPTVET-----ADIAYAVVEIDGASAVGGIETQYLDDSEVDLEVYDAGV 363

Qy        274 AANAA--ESATKD-QLTQE----AFKNPENQKV---NIDANGNAIPSGELXDDIVEQIAQ 323
              :| ||  : || |  || |       :|  | |    :| || | |          : :|
Db        364 SATAAVGQDATNDITLTIEEGGTTLSSPTGQYVVGSEVDINGTATSS--------DSVAI 415

Qy        324 QAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVT 383
               :: |:    | :|    :   :      :|:: ||   | | || | : |    ||
Db        416 YVRDDGD---WQLLEIGGDNEISVDSDDTFEEEDIALSGLSGDG-SSILSLTGTYRIGVI 471

Qy        384 TALHR---RNQPAEQTTTTTTHT----------VVQQQTGGIPQHKVALMPQERRRFSD 429
               |      :  : : ||:  |              |  ||        : | |
Db        472 DASDADVGGDGSVDDSLTTSEFTSGVSSSNSIRVTDQALTGQFTTINGQVAPVE------ 525

Qy        430 RRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFS 489
                |:|     :  :: |:  : :  | |           | ||: |   :  ::
Db        526 ----TGTVDINGTASGANSVLVIFVDERGNVN----------YQEVSVDSDGTYDEDDI 570

Qy        490 GSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSG 529
               |   ||:          |  :|:      :| | | ||
Db        571 TVGLTQGRV----------TAHILSVGRDSAIGDGSLPSG 600


RESULT 13
ZEST_DROVI
ID   ZEST_DROVI     STANDARD;      PRT;   618 AA.
AC   Q24762;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Regulatory protein zeste.
GN   Z.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7244;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92123185; PubMed=1732733;
RA   Chen J.D., Chan C.S., Pirrotta V.;
RT   "Conserved DNA binding and self-association domains of the Drosophila
RT   zeste protein.";
RL   Mol. Cell. Biol. 12:598-608(1992).
CC   -!- FUNCTION: INVOLVED IN TRANSVECTION PHENOMENA (= SYNAPSIS-DEPENDENT
CC       GENE EXPRESSION), WHERE THE SYNAPTIC PAIRING OF CHROMOSOMES
CC       CARRYING GENES WITH WHICH ZESTE INTERACTS INFLUENCES THE
CC       EXPRESSION OF THESE GENES. ZESTE BINDS TO DNA AND STIMULATES
CC       TRANSCRIPTION FROM A NEARBY PROMOTER (BY SIMILARITY).
CC   -!- SUBUNIT: SELF-ASSOCIATE FORMING COMPLEXS OF SEVERAL HUNDRED
CC       MONOMERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M76700; AAA29052.1; -.
DR   FlyBase; FBgn0013149; Dvir\z.
KW   DNA-binding; Transcription regulation; Nuclear protein.
FT   DNA_BIND     66    146       SPECIFIC, WITH ZESTE LOCUS.
FT   DOMAIN      171    468       GLN/ALA-RICH (OPA-REPEAT INVOLVED IN
FT                                TRANSCRIPTIONAL ACTIVATION OR REPRESSION
FT                                AT DIFFERENT TARGET LOCI) (POTENTIAL).
SQ   SEQUENCE   618 AA;  68102 MW;  3127180B10AFDC0B CRC64;

  Query Match             4.7%;  Score 131.5;  DB 1;  Length 618;
  Best Local Similarity   22.6%;  Pred. No. 1.4;
  Matches   66;  Conservative   57;  Mismatches  116;  Indels   53;  Gaps   15;

Qy        217 SETHTSTTNSSVRSDPKFWVSV----------GAIAAGLAGLAATGIAQALALTPEPDDP 266
              |  | |    :::   |||::           |: : | |||   |   | |   :
Db        120 SHPHVSAMR-NIKQIQKFWLNSRLRKQYPYRDGSASGGSAGLGKVGTVSASAQQQQQQQQ 178

Qy        267 TTTDPDQAA--NAAESATKDQLTQEAFK-NPENQKVN-IDANGNAIPSGELXDDIVEQIA 322
                        |    :: :   : |:: || : ||: :  :  |: :|| :   |: || :|  |
Db        179 QQQQQQQQQHHDSVKVEPEYQISPEASEHNPQGEPFDEIEMDGNDV--SEMEDDPLE--A 234

Qy        323 QQAKEAGEVARQQ-AVESNAQAQQRYEDQHA-RRQEELQLSSGIGYGLSSAL----IVAG 376
              || ::  :  :|| | :: |:|||: : | |     ::||:|: :    :| |     |
Db        235 QQQQQQQQQQQQQAAAQAQAEAQQQQQQQSAVAEMQKLQVSAAVAAANASMLNTHRINVD 294

Qy        377 GIGAGVTT---ALH----RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSD 429
               | |    |     ||    | ::   |    |  | |          | :  ||
Db        295 SISAEKLTLNDLLHFKPARHDEIILQIKHPTDATATQ-------IHTIPAQPQ------- 340

Qy        430 RRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPG 481
                  | ::|:       : :::    :|:   : :|: :| ::|    :    ||
Db        341 ----QHTMATITAGGYNQQII---SEIKPQQITLAQYQAQQHQQAQARLRPG 385


RESULT 14
MAZ4_SCHCO
ID   MAZ4_SCHCO     STANDARD;      PRT;   940 AA.
```

```
AC   P37938;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mating-type protein A-alpha Z4.
OS   Schizophyllum commune (Bracket fungus).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Agaricales; Schizophyllaceae; Schizophyllum.
OX   NCBI_TaxID=5334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 340.81 / UVM 4-40;
RX   MEDLINE=92357793; PubMed=1353886;
RA   Stankis M.M., Specht C.A., Yang H., Giasson L., Ullrich R.C.,
RA   Novotny C.P.;
RT   "The A alpha mating locus of Schizophyllum commune encodes two
RT   dissimilar multiallelic homeodomain proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:7169-7173(1992).
CC   -!- FUNCTION: SPECIFIES A-ALPHA-4 MATING-TYPE. MAY REGULATE THE
CC       EXPRESSION OF GENES SPECIFIC TO THE HOMOKARYOTIC CELL TYPE.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED CONSTITUTIVELY IN HOMOKARYONS.
CC   -!- SIMILARITY: BELONGS TO THE TALE/M-ATYP HOMEOBOX FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M97181; AAB01372.1; -.
DR   PIR; D37271; D37271.
DR   InterPro; IPR001356; Homeobox.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; FALSE_NEG.
DR   PROSITE; PS50071; HOMEOBOX_2; FALSE_NEG.
KW   Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT   DNA_BIND    110    182       HOMEOBOX (TALE-TYPE).
FT   DOMAIN      370    402       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      436    469       ASP/GLU-RICH (ACIDIC).
SQ   SEQUENCE   940 AA;  101856 MW;  4B99CBAEDB39621E CRC64;

  Query Match             4.7%;  Score 130.5;  DB 1;  Length 940;
  Best Local Similarity   22.5%;  Pred. No. 2.7;
  Matches  113;  Conservative   63;  Mismatches  205;  Indels  121;  Gaps   21;

Qy         23 PSQTDGAARGGTGHLISSTGALGSRSLFSP--------LRNSMADSVDSR--DIPGLPTN 72
              | :  || ||       |  |   | :|||        :| |:     |     |  | :
Db        470 PVKIAGAKRGRNDDEEVSPLAKKPR-IFSPPVRPRPQAIRVSLPSPAPSSRGSTPTSPVS 528

Qy         73 PS----RLAAATSETCLLGGFEV-LHDKGPLDILNTQIGPSAFRVEVQADGTH-AAIGEK 126
              ||     | | |||   ||      :   :|   ::    : | :   | :  ||     :  :
Db        529 PSPKAKRPAQATS---LLASHPMKKREKLQEELRKAGLAPPSAPVLMGPDGVPLGTVRSR 585

Qy        127 NGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTRILAKLD 186
               :   ||   ||   |    : :  |     | :||    :| |    |  |: || |
Db        586 SSPSVS---SPPSVSVSLPLPSRG----VPSGGIKVTGDPTPWVNWDL-EAHT------- 630

Qy        187 PDNHGGRQPKD------------VD-----------TRSVGVGSASGIDDGVVSETHTST 223
                     : |:|            ||           |||  : | |         :  ||
Db        631 ------QAPRDLTAATKSSAGCSVDAVPLPGKSRSLTRSPSISSIS---------SACST 675

Qy        224 TNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATK 283
              ::|    :|  | |:                    :  |  : |  |||:  ||  |      :  : |:::
Db        676 SSSGSDTDSLFSVT----------------SDATDIT-EPDEATTADETTTQSTSASSSR 718

Qy        284 DQLTQEAFKNPENQKVNID------------ANGNAIPSGEL---------XDDIVEQIA 322
              |    :|:    |  :       |                 |:|   || |         |  |  :|
Db        719 DTTSQQKRMPPLSIDPRFDPALWSKYDLSPPADGRLHPSDGLRPSAFVPTKLDVRVANLA 778

Qy        323 QQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAG---GIG 379
```

```
                                |        :::     ::  |          ||                |  |   ::::   |    |
Db        779 QNPARHWSASKRSPTRASHAAAPIVSYHHATGSIASPAQVAFGEGQLTSVLATGQKAGNA 838

Qy        380 AGVTTALHRRNQPAEQTTTTTHTVVQQQTGGIPQ-----HKVALMPQERRRFSDRRDSQ 434
                    || : ||   |   |   | |:    :   |      : |:            |  |    |:   ||: :||: :
Db        839 RRRTTPVQRRVTPKAQETSEPSSLVDGILSSGLADVCREAPKPAKAPKNDRRYLERRERR 898

Qy        435 GSVASTHWSDSSSEVVNPYAEV 456
                    | :|       ||:   |      ||:
Db        899 LSKSSP--VDSADTVRTRLAEI 918


RESULT 15
ALS1_CANAL
ID   ALS1_CANAL     STANDARD;      PRT;  1260 AA.
AC   P46590;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin-like protein 1 precursor.
GN   ALS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11651 / B792;
RX   MEDLINE=95272392; PubMed=7752895;
RA   Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT   "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT   sexual agglutinin separated by a repeating motif.";
RL   Mol. Microbiol. 15:39-54(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST SAG1.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L25902; AAC41649.2; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1260       AGGLUTININ-LIKE PROTEIN 1.
FT   DOMAIN      433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT      433    468       1-1.
FT   REPEAT      469    504       1-2.
FT   REPEAT      505    540       1-3.
FT   REPEAT      541    576       1-4.
FT   REPEAT      577    612       1-5.
FT   REPEAT      613    648       1-6.
FT   REPEAT      649    684       1-7.
FT   REPEAT      685    720       1-8.
FT   REPEAT      721    756       1-9.
FT   REPEAT      757    792       1-10.
FT   DOMAIN      983   1152       2 X 26 AA APPROXIMATE REPEATS.
FT   REPEAT      983   1043       2-1.
FT   REPEAT     1092   1152       2-2.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      450    455       POLY-THR.
FT   DOMAIN      486    491       POLY-THR.
FT   DOMAIN      522    527       POLY-THR.
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
```

```
FT   DOMAIN       738    743       POLY-THR.
FT   DOMAIN       774    779       POLY-THR.
FT   DOMAIN       874    877       POLY-SER.
FT   CARBOHYD     471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     973    973       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1045   1045       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;

  Query Match             4.6%;  Score 130;  DB 1;  Length 1260;
  Best Local Similarity   20.6%;  Pred. No. 4.2;
  Matches   73;  Conservative   48;  Mismatches  167;  Indels   66;  Gaps   12;

Qy         16 IPPAPPLP---SQTDGAARGGTGHLISS-------TGALGSRSLFSPLRNSMADSVDSRD 65
              :| |  :    | ||| :   : ::  |       |     ::  :|| :|:
Db        895 VPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSS 954

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRV--EVQADGTHAAI 123
              || :  : |::   ::     |  :       ::|: |    |: :|   | :  |
Db        955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTT----SSTKVLPPVVSSNTDLTS 1010

Qy        124 GEKNGLEVSVTLSPQEWSSLQSIDTE---GKNRFVFTGGRGGSGHPMVTVA-SDIAEART 179
               |  |   |||    |  : | |    | |     |    | :|: ||| | :| |
Db       1011 EPTNTREQPTTLSTTSNSITEDITTSQPTGDN-----GDNTSSTNPVPTVATSTLASAS- 1064

Qy        180 RILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVG 239
                : ||  |      :: |  :  : | : |: : |    | :|    |   || |
Db       1065 ------EEDNKSGSH----ESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSG 1114

Qy        240 AIAAGLAGLAATGIAQALALTPEPDD----PTTTDPDQAANAAESATKDQLTQEAFKNPE 295
                          :| || |    |||       |:   ||     :   |:|
Db       1115 T----------------DVTTEPTDTREQPTTLSTTSKTNSESVATTQATNENGGKSPS 1157

Qy        296 NQ---------KVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESN 340
                 :  ||   : || :    |  :      :  |    :: ||
Db       1158 TDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSN 1211


Search completed: December 16, 2003, 08:53:50
Job time : 24.306 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 26.2608 Seconds
(without alignments)
2010.470 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1528.5	54.6	558	2	A98199	translocated intim
2	1528.5	54.6	558	2	E86045	probable transloca
3	151.5	5.4	2793	2	B90784	hypothetical prote
4	150.5	5.4	1608	2	A28182	hemolysin A - Serr
5	150	5.4	1005	2	C71513	hypothetical prote
6	149.5	5.3	2806	2	D85644	hypothetical prote
7	149	5.3	2232	2	T34434	hypothetical prote
8	146	5.2	1829	2	T24583	hypothetical prote
9	144.5	5.2	1296	2	T13936	collar protein iso
10	144	5.1	883	2	S04722	puff 74E protein -
11	144	5.1	1589	2	T42233	submaxillary mucin
12	143.5	5.1	2551	2	B98047	hypothetical prote
13	141.5	5.1	647	2	T39141	hypothetical prote
14	141	5.0	2535	2	AC0304	probable hemolysin
15	140	5.0	1152	2	AE1852	hypothetical prote
16	140	5.0	2132	1	A55182	aggrecan precursor
17	137	4.9	1532	2	A26039	IgA-specific metal
18	137	4.9	2468	2	A83412	hypothetical prote
19	136.5	4.9	770	2	T51024	related to C2H2 zi
20	136.5	4.9	1829	2	T35681	probable sensory h
21	136	4.9	674	2	B82117	probable flagellar
22	136	4.9	1569	2	A65044	hypothetical prote
23	135.5	4.8	2124	2	A28452	proteoglycan core
24	135	4.8	1085	2	S66149	gene pipsqueak pro
25	134.5	4.8	3295	2	AE0074	probable adhesin Y
26	134.5	4.8	13288	2	T03099	mucin, submaxillar
27	134	4.8	3013	2	AB0480	probable invasin Y
28	133.5	4.8	1975	2	B81192	hemagglutinin/hemo
29	133.5	4.8	1995	2	G81044	hemagglutinin/hemo
30	133.5	4.8	2174	2	E95965	hypothetical glyci
31	133	4.8	1571	2	C91068	hypothetical prote
32	133	4.8	3442	2	E82589	hemagglutinin-like
33	133	4.8	3455	2	B82519	hemagglutinin-like
34	132.5	4.7	812	2	T48016	probable zinc-fing
35	132.5	4.7	1026	2	A48995	paracrystalline su
36	132.5	4.7	1073	2	C87374	S-layer protein Rs
37	132.5	4.7	1630	2	A53577	ascites sialoglyco
38	132	4.7	666	2	G82111	flagellar hook-ass
39	132	4.7	827	2	A37849	S-layer protein -
40	132	4.7	1528	2	D85912	hypothetical prote
41	132	4.7	5188	2	B85547	probable RTX famil
42	132	4.7	5291	2	F90696	hypothetical prote
43	131.5	4.7	635	2	F75477	hypothetical prote
44	131.5	4.7	1804	2	H96597	hypothetical prote
45	131	4.7	2703	1	A24420	notch protein - fr

ALIGNMENTS

RESULT 1
A98199
translocated intimin receptor Tir [imported] - Escherichia coli (strain O157:H7, substra
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: A98199
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A98199
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-558 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37984.1; PID:g13364036; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4561

Query Match 54.6%; Score 1528.5; DB 2; Length 558;
Best Local Similarity 56.2%; Pred. No. 6.5e-88;
Matches 327; Conservative 60; Mismatches 138; Indels 57; Gaps 11;

```
Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| | |: ||||||||||||||  || | ||:||| ||||:||:|:|||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
               |:|  |:|||| || ||||  ||  |  ||||||| |||| || ||| | |||| | ||
Db         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
               | |:|::||:| || || ||:: ||||| |||::|||||||||:|| ||||||||  |||
Db        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176

Qy        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230
               |||  |:|  |          | | | : |:      : |    :|| |||: ||:||
Db        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227

Qy        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290
              ||| |:::| :| || ||||||| |||||||||| ||||||| ||:| |:||:||||:||
Db        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287

Qy        291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350
              |:||:|||||||  |||||||  | ||:|  | :||| ||| |:|||:|:|||||::|::|
Db        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347

Qy        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403
              |:|||||::||| ||||| |||: |||| || ||||:||| ||||||||         |
Db        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSART 407
```

```
Qy        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSS-SEVVNPYAEVGGARNS 462
              |  :     |       |         || |  | :|| :  ||   | ||||:|
Db        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGTVQNPYADV-----K 462

Qy        463 LSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVT---------------GRLIGTPGQGIQ 507
              | |       :|      : :||   :  |                  ||:| |  |||
Db        463 TSLHD------SQVPTSNSNTSVQNMGNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQ 516

Qy        508 STYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549
              |||| || |||||  ||||| |  :||:: |  | ||  |||
Db        517 STYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 558


RESULT 2
E86045
probable translocated intimin receptor protein tir [imported] - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E86045
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E86045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-558 <STO>
A;Cross-references: GB:AE005174; NID:g12518449; PIDN:AAG58825.1; GSPDB:GN00145; UWGP:Z51
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: tir

  Query Match             54.6%;  Score 1528.5;  DB 2;  Length 558;
  Best Local Similarity   56.2%;  Pred. No. 6.5e-88;
  Matches  327;  Conservative   60;  Mismatches  138;  Indels   57;  Gaps   11;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| | |: |||||||||||||||  || | ||:||| ||||:||:|:|||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
               |:|  |:|||| || ||||  ||  |  ||||||| |||| || ||| | |||| | ||
Db         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
              | |:|::||:| || || ||:: ||||| |||::|||||||||:|| |||||||| |||
Db        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176

Qy        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230
              |||  |:|   |           | | | : |:      : |    :|| |||: ||:||
Db        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227

Qy        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290
              |||  |:::| :| || ||||||| |||||||||| ||||||| ||:| |:||:||||:||
Db        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287

Qy        291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350
              |:||:|||||||  |||||| | ||:|  | :||| ||| |:|||:|:|||||::|:::|
Db        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347

Qy        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403
              |:|||||::||| ||||| |||: ||||  || ||||:||| ||||||||         |
Db        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTTSART 407

Qy        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSS-SEVVNPYAEVGGARNS 462
              |  :     |       |         || |  | :|| :  ||   | ||||:|
Db        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSSIGTVQNPYADV-----K 462

Qy        463 LSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVT---------------GRLIGTPGQGIQ 507
              | |       :|      : :||   :  |                  ||:| |  |||
Db        463 TSLHD------SQVPTSNSNTSVQNMGNTDSVVYSTIQHPPRDTTDNGARLLGNPSAGIQ 516

Qy        508 STYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549
              |||| || |||||  ||||| |  :||:: |  | ||  |||
Db        517 STYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 558


RESULT 3
B90784
hypothetical protein ECs1242 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90784
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90784
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2793 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34665.1; PID:g13360702; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1242

  Query Match              5.4%;  Score 151.5;  DB 2;  Length 2793;
  Best Local Similarity   21.4%;  Pred. No. 0.61;
  Matches  115;  Conservative   87;  Mismatches  222;  Indels  113;  Gaps   27;

Qy         22 LPSQTDGAARG--GTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAA 79
              ||:  | |:  |||  :      ||   : ||::   |  ::    |||         |
Db         55 LPTMGKGFAQSVRGTGEM---ARGLGDAMIQSPVKTG-ARILNEFSRMGLP------GVA 104

Qy         80 TSETCLLGGFEVLHDKGPLDILNT-QIGPSAFRVEV----QADGTHAAIGEKNGLE-VSV 133
              | :    ||      :|  ::::|   | :|    |    :| |   : | |   | ::
Db        105 TVQDIFAGG-----SRGADEVIDTLPDGKNAVTDTVGKGLKATGKAVSDGAKATDEWLTG 159

Qy        134 TLSPQEWSSLQSIDTEGKNRFVFTGGRG----GSGHP-MVT--VASDIAEARTR------ 180
               :||    :| :  ||| |       :|    |: | ||   ||  : :   |
Db        160 KMSPGAVRALNTPMTEGYNDSAVWVAKGVNLIGALVPDMVAGGVARKVGDVTLRKMLTAG 219

Qy        181 -----ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFW 235
                  | | : |:       : || :   :  |     |::: :  |    |
Db        220 LEKKYIAAGMQPERATALAAEAVDKKMPDLFQA-----GLITHSTVSAQGQS-------- 266

Qy        236 VSVGAIAAGLAGLAA--TGIAQA-------LALTPEPDDPTTTDPDQAANAAESATKDQL 286
                  |:||  | | |  : :||:       |::  :|     ||  |  : |:    |::
Db        267 ----AMAAADAVLNADYSELAQSPKFQQTFLSIDADPQHAQLTD-RQKMDLAKERVADEV 321

Qy        287 TQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKE--AGEVARQQAVESNAQAQ 344
              :   :||   ||  |    :  :| : :    |:  |          | |: :
Db        322 RAQLATDPELLAVN--AMAAKLGDAQLFNLVTRGTAKTVKSGIVRNATEQGAINAAQGGY 379

Qy        345 QRYEDQHARRQEELQLSSGIGY----GLSSALIVAGGIGA------GVTTALHRRNQPAE 394
              ||::  | |:     ::|:|     |:: | |     ||      |       | | ||
Db        380 SRYQENTALRE-----TAGMGVSPWEGVADATIEGAAFGAAMGAPFGAVAGYRGRRQAAE 434

Qy        395 QTTTTTTHTVVQQQTGGIPQ-HKVALMPQER-----------RRFSDR-RDSQGSVAST 440
              :|      ||  ||    ||    |  : |:|           :::|    ::|  ::
Db        435 ETAMRDAETV--QQDDAAPQPESVDPVAQQRESMQGMNREQLLEQYADADMATEGDASAA 492

Qy        441 HWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGR 497
              |  :::|:::           | |       : : :|: | |   ::: :       ||
Db        493 HRREAASQLL----------NELDEQTKRQAVMNELKAKPRSELLEEYRRLSQKEGR 539


RESULT 4
A28182
hemolysin A - Serratia marcescens
C;Species: Serratia marcescens
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Oct-1999
C;Accession: A28182
R;Poole, K.; Schiebel, E.; Braun, V.
```

```
J. Bacteriol. 170, 3177-3188, 1988
A;Title: Molecular characterization of the hemolysin determinant of Serratia marcescens.
A;Reference number: A28182; MUID:88257037; PMID:3290200
A;Accession: A28182
A;Molecule type: DNA
A;Residues: 1-1608 <POO>
A;Cross-references: GB:M22618; NID:g340726; PIDN:AAA50323.1; PID:g556420

  Query Match             5.4%;  Score 150.5;  DB 2;  Length 1608;
  Best Local Similarity   20.9%;  Pred. No. 0.32;
  Matches  122;  Conservative   84;  Mismatches  236;  Indels  141;  Gaps   25;

Qy         33 GTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSR--------LAAATSETC 84
              |:  : : |  |::   |     ::  |:|: :  |:  |  :          |    :|
Db       1008 GSDLTVDAKGEGGTQRSNSSASQAVTGSIDAAN--GINVNVKKDAIYQGTALNGGRGKTA 1065

Qy         85 LLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA----------------DGTH-------- 120
              :  | ::  |:        ::   | | |:  |                 |||
Db       1066 VNAGGDIRLDQASDKQSESR---SGFNVKASAKGGFTADSKNFGAGFGGGTHNGESSSST 1122

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEA-RT 179
              | :|  :| :  | |       :||  | : :     : |              ||   ||: :|
Db       1123 AQVGNISG-QQGVELKAGRDLTLQGTDVKSQGDVSLSAGN--------KVALQAAESTQT 1173

Qy        180 RILAKLDPDNHGGRQPKDVDTRSVGVGSASG-IDDGVVSETHTSTTNSSVRSDPKFWVSV 238
              |  :||  :    |     |   :: |  || |  |   |:|: |    ::: || |  :|
Db       1174 RKESKLSGNIDLGAGSSDSKEKTGGNLSAGGAFDIAKVNESATERQGATIASDGKVTLSA 1233

Qy        239 ---GAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAA---ESATKDQLTQEAFK 292
                 |  | | |   :| : ||                 :| |     |||  :|
Db       1234 NGKGDDALHLQGAKVSGGSAAL---------------EAKNGGILLESAKNEQ------- 1271

Qy        293 NPEN------------QKVNIDANGNAIP-----SGELXDDIVEQIAQQAK--------E 327
              : :|             | | || |   |     :  |   :  : || |
Db       1272 HKDNWSLGIKANAKGGQTFNKDAGGKVDPNTGKDTHTLGAGLKVGVEQQDKTTHANTGIT 1331

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
              ||:|     ::   |  | :       : :|   |    :  |   |  |:   |    |
Db       1332 AGDVTLNSGKDTRL-AGARVD------ADSVQGKVGGDLHVESRKDVENGVKVDVDAGLS 1384

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQ------HKVALMPQER-----RRFSDRRDSQGS 436
               | |    |:  :     :  | : :      :|||    ::     ||   ::|: |:
Db       1385 HSNDPGSSITSKLSKVGTPRYAGKVKEKLEAGVNKVADATTDKYNSVARRLDPQQDTTGA 1444

Qy        437 VASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTG 496
              |:   :| :  :|  |   |          :|:  ::|  |   | :|  : :|     |
Db       1445 VS---FSKAEGKVTLPATPAG--------EKPQGPLWDRGARTVGGAVKDSITGPAGRQG 1493

Qy        497 RL------IGTPGQGIQSTYALLANSGGLRLGMGGLT--SGGE 531
               |      :     | ||    :|  |: | :|| |  :|||
Db       1494 HLKVNADVVNNNAVGEQSA---IAGKNGVALQVGGQTQLTGGE 1533


RESULT 5
C71513
hypothetical protein CT456 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: C71513
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: C71513
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1005 <ARN>
A;Cross-references: GB:AE001319; GB:AE001273; NID:g3328881; PIDN:AAC68056.1; PID:g332888
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT456

  Query Match             5.4%;  Score 150;  DB 2;  Length 1005;
  Best Local Similarity   22.8%;  Pred. No. 0.18;
  Matches  145;  Conservative   65;  Mismatches  234;  Indels  192;  Gaps   33;

Qy          4 GNLGNNVNGNHLIPPAPPL-PSQTDGAARGGT------GHLISSTGALGSRSLFSPLR-- 54
              ||| | |  |:          |  :  |: | |      || :  |     : |: :
Db        431 GNLPNTVIINNKFKTCVAYGPWNSQEASSGYTPSAWRRGHRVDFGGIFEKANDFNKINWG 490

Qy         55 -----NSMADSVD-SRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPS 108
                  :|  | :  | : ||    |   |||  | |              : |:|
Db        491 TQAGPSSEDDGISFSNETPG--AGP---AAAPSPT----------PSSIPIIN------ 528

Qy        109 AFRVEVQADGTHAAIGEKNGLEVSVT-LSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPM 167
                | |   ||:  ||: |   |: |  :|   |:  | ||
Db        529 ---VNVNVGGTNVNIGDTN---VNTTNTTPTTQSTDASTDT------------------ 563

Qy        168 VTVASDIAEARTRILAKLDPDNHGGRQPKDVDTR---SVGVG----------SASGIDDG 214
                |||  :  |             |  |::|    | | |                :|| | |
Db        564 ----SDIDDINT-----------NNQTDDINTTDKDSDGAGGVNGDISETESSSGDDSG 607

Qy        215 VVSETHTSTTNSSVRSDPKFWVSV--------------------GAIAAG------LAG 247
              || :  |  |:|| :|       :                    | : |        ::
Db        608 SVSSSE-SDKNASVGNDGPAMKDILSAVRKHLDVVYPGENGGSTEGPLPANQTLGDVISD 666

Qy        248 LAATGIAQALALTPE----PDDPTTTDPDQAANAAESAT--------KDQLTQEA----- 290
              :  | ||   |:       ||||||       | ||  |           | ::  |
Db        667 VENKGSAQDTKLSGNTGAGDDDPTTT--AAVGNGAEEITLSDTDSGIGDDVSDTASSSGD 724

Qy        291 ----FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQR 346
                  :| ::     | ||  |||    ||: : :   :       :  |   || |
Db        725 ESGGVSSPSSESNKNTAVGNDGPSGL---DILAAVRKHLDKVYPGDNGGSTEGPLQANQT 781

Qy        347 YED-----QHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT 401
               |      :    :|  :|     | :|:   ||| |:  |  |      |   |||  |
Db        782 LGDIVQDMETTGTSQETVVSP--WKGSTSSTESAGGSGSVQT--LLPSPPPTPSTTTLRT 837

Qy        402 HT----------------VVQQQTGG-IPQHKV--ALMPQERRRFSDRRDSQGSVASTHW 442
               |                ::    || ||      |:|:  |       |:|| : ||
Db        838 GTGATTTSLMMGGPIKADIITTGGGGRIPGGGTLEKLLPRIRAHLDISFDAQGDLVSTEE 897

Qy        443 SDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTP 502
               | :|| : :  |:|  |:         :     || :  |  :|:|       |
Db        898 PQLGS-IVNKFRQETGSRGILA-------FVESAPGKPGSA--QVLTGTG-------GDK 940

Qy        503 GQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVN 538
               |   |: |:    | :  |   |   |:   | ||
Db        941 GNLFQAAAAVTQALGNV-AGKVNLAIQGQKLSSLVN 975


RESULT 6
D85644
hypothetical protein Z1495 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85644
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85644
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2806 <STO>
A;Cross-references: GB:AE005174; NID:g12514354; PIDN:AAG55616.1; GSPDB:GN00145; UWGP:Z14
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1495

  Query Match             5.3%;  Score 149.5;  DB 2;  Length 2806;
  Best Local Similarity   20.8%;  Pred. No. 0.82;
  Matches  113;  Conservative   88;  Mismatches  218;  Indels  123;  Gaps   27;

Qy         22 LPSQTDGAARG--GTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAA 79
```

```
Db        68 LPTMGKGFAQSVRGTGEM---ARGLGDAMIQSPVKTG-ARILNEFSRMGLP------GVA 117

Qy        80 TSETCLLGGFEVLHDKGPLDILNTQ----------IGPSAFRVEVQADGTHAAIGEKNGL 129

Db       118 TVQDIFAGG-----SRGADEVIDTLPDGKNAVTDIVGKG-----LKATGKAVSDGAKATD 167

Qy       130 E-VSVTLSPQEWSSLQSIDTEGKNRFVFTGGRG----GSGHP-MVT--VASDIAEARTR- 180

Db       168 EWLTGKMSPGAVRALNTPMTEGYNDSAVWVAKGVNLIGALVPDMVAGGVARKVGDVTLRK 227

Qy       181 ----------ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRS 230

Db       228 MLTAGLEKKYIAAGMQPERATALAAEAVDKKMPDLFQA-----GLITHSTVSAQGQS--- 279

Qy       231 DPKFWVSVGAIAAGLAGLAA--TGIAQA-------LALTPEPDDPTTTDPDQAANAAESA 281

Db       280 ---------AMAAADAVLNADYSELAQSPKFQQTFLSIDADPQHAQLTD-RQKMDLAKER 329

Qy       282 TKDQLTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKE--AGEVARQQAVES 339

Db       330 VADEVRAQLATDPELLAVN--AMAAKLGDAQLFNLVTRGTAKTVKSGIVRNATEQGAINA 387

Qy       340 NAQAQQRYEDQHARRQEELQLSSGIGY----GLSSALIVAGGIGA------GVTTALHRR 389

Db       388 AQGGYSRYQENTALRE-----TAGMGVSPWEGVADATIEGAAFGAAMGAPFGAVAGYRGR 442

Qy       390 NQPAEQTTTTTTHTVVQQQTGGIPQ-HKVALMPQER-----------RRFSDR-RDSQG 435

Db       443 RQAAEETAMRDAETV--QQDDAAPQPESVDPVAQQRESMQGMNREQLLEQYADADMATEG 500

Qy       436 SVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVT 495

Db       501 DASAAHRREAASQLL----------NELDEQTKRQAVMNELKAKPRSELLEEYRRLSQKE 550

Qy       496 GR 497

Db       551 GR 552
```

RESULT 7
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

  Query Match 5.3%; Score 149; DB 2; Length 2232;
  Best Local Similarity 19.9%; Pred. No. 0.64;
  Matches 112; Conservative 71; Mismatches 238; Indels 142; Gaps 19;

```
Qy        21 PLPSQ--TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPG----LPTNPS 74

Db       899 PSPSQSTTIGSTQGSTSPGISTT-----------SEEMTSQGSTQTPGSTGSTVTQPS 945

Qy        75 RLAAATS--ETCLLGGFEVLHDKGPLDILNT--------------QIGPSAFRVEVQAD 117

Db       946 TVSDSTSSGSTVTVGSTEGSSSPIPSTSQNTNPSTSSGSSMSTQTPQSSQSTSPVESSTS 1005
```

```
Qy       118 GTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEA 177

Db      1006 GATSSSGSPGTTLTSISPSPSPSSTI-------------GSSQGSTSPVV---STISQG 1048

Qy       178 RTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVS 237

Db      1049 STE-----TPGSTGSTVTKP----STVSGSASSGSTATMGSTEASSTSGGSSTSPNPSQS 1099

Qy       238 VGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA--AESATKDQLTQEAFKNPE 295

Db      1100 TSPSTSGATSSPGSSGTTLTSISPSPSQSSTIGSSQGSTSPVVSTTSGDMTSQGSTQIPG 1159

Qy       296 NQKVNI----DANGNAIPSGELXDDIVEQIAQQA---KEAGEVARQQAVESNAQAQQRYE 348

Db      1160 STGSTVTQPSTGSGSTSTSGEITSQGSTQTPRSSLSTSPAISTSTQQSVSTNSP------ 1213

Qy       349 DQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQ 408

Db      1214 ----------------------------GSTVTQPSTVRGSTSSGSTVTTGSTEGSST 1243

Qy       409 TGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQP 468

Db      1244 SGS-------------------SSATSLSSSSPVPSTSQSPNP----STSGSSTPTPNP 1279

Qy       469 EEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMG---G 525

Db      1280 SQSTSPVVSTTTGEMTSHGSTQTPSTIGSTVTQP-----STVS-GSNSSGSTVTIGSSEA 1333

Qy       526 LTSGG--ETAVSSVNAAPTPGPV 546

Db      1334 STSGSSFKTSPSSISPVPTSSPI 1356
```

RESULT 8
T24583
hypothetical protein T06D8.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24583
R;Palmer, S.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19909
A;Accession: T24583
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1829 <WIL>
A;Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1
A;Experimental source: clone T06D8
C;Genetics:
A;Gene: CESP:T06D8.1
A;Map position: 2
A;Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

  Query Match 5.2%; Score 146; DB 2; Length 1829;
  Best Local Similarity 21.7%; Pred. No. 0.74;
  Matches 118; Conservative 63; Mismatches 256; Indels 108; Gaps 19;

```
Qy        68 GLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTHAAIGEKN 127

Db       172 GEETTTSAVTEASSEA----------------TTTPAGTEASGEET----TTSAVTEGS 210

Qy       128 GLEVSV-----TLSPQEWSSLQSIDTE-GKNRFVFTGGRGGSGHPMVTVASDIAEARTRI 181

Db       211 GEETTVVAVVESSGEEPASSSTSIPTELSKNDQV----TEASGEETITAAATEASEETTT 266

Qy       182 LAKLDPDNHGGRQPKDVDTRSVGVGSASG--------------IDDGVVSETHTSTTNSS 227

Db       267 SAVTEGSGE--------DTTVVAVVELSGEQPASSSTSIPTELSKDDQVTEASGEETTTA 318

Qy       228 VRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDP------DQAANAAESA 281

Db       319 AATEASEETTTSAVTEG-SGEETTVVAVVESSGEEPASSSTSIPTELSKDDQVTEASGEE 377
```

```
Qy        282 TKDQLTQEAFKNPENQKVN------------IDANG-------NAIPSGELX--DDIVEQ 320
Db        378 TTTAAATEASEETTTSAVTEGSGEDTTVVAVVESSGEQPASSSTSIPT-ELSKDDQVTEA 436

Qy        321 IAQQAKEAGEVARQQAVESNAQAQQRYEDQ------HARRQEELQLSSGIGYGLSSALIV 374
Db        437 SGEETTTAAATEASEETTTSAVTEGSGEDTTVVAVVESSGEQPASSSTSIPTELSKDDQV 496

Qy        375 AGGIGAGVTTALHRRNQPAEQTTTT---------TTHTVVQQQTGGIPQHKVALMPQERR 425
Db        497 TEASGEETTTA--AATEASEETTTSAVTEGSGEETTVVAVVESSGEEPASSSTSIPTELS 554

Qy        426 RFSDRRDSQGSVAST-HWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSV 484
Db        555 KDDKVTEASGEETTTAAATDASSEETTTSAVTEGSGEETTVVAVVES-SDEEPASSSTSI 613

Qy        485 IQNFS-------GSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSV 537
Db        614 PTELSKDDQVTEASGEETTTAAATEASEETTTSAVTEGSGEETTVVAVVESSGEEPASSS 673

Qy        538 NAAPT 542
Db        674 TSIPT 678
```

RESULT 9
T13936
collar protein isoform C - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T13936
R;McGinnis, N.; Ragnhildstveit, E.; Veraksa, A.; McGinnis, W.
Development 125, 4553-4564, 1998
A;Title: A cap 'n' collar protein isoform contains a selective Hox repressor function.
A;Reference number: Z17817; MUID:98453413; PMID:9778513
A;Accession: T13936
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1296 <MCG>
A;Cross-references: EMBL:AF070064; NID:g3859888; PID:g3859889; PIDN:AAC72898.1
C;Genetics:
A;Gene: cnc
A;Cross-references: FlyBase:FBgn0000338
A;Map position: 3
C;Keywords: leucine zipper

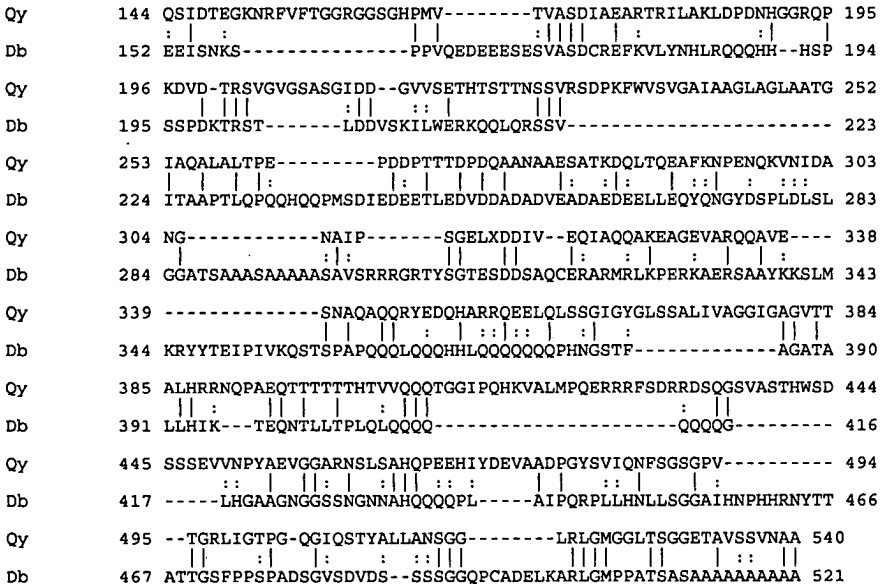
  Query Match 5.2%; Score 144.5; DB 2; Length 1296;
  Best Local Similarity 19.9%; Pred. No. 0.56;
  Matches 149; Conservative 68; Mismatches 213; Indels 319; Gaps 31;

```
Qy          8 NNVNGNHLIPPAPPLPSQ-TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDI 66
Db         24 HNTTGNSSVQTAALQDVQSTSAAATGAT--MVVGTG---------------------- 57

Qy         67 PGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTHAAIGEK 126
Db         58 -GAPTSSGQTSGSA-----LG--EIHIDTASLDPGNANHSPLHPTSELDTFLTPHALQDQ 109

Qy        127 NGL-----------EVSVTLSPQEWSSLQSIDTEGKN--------RFVFTGGRGGSGHP 166
Db        110 RSIWEQNLADLYDYNDLSLQTSPYANLPLKDGQPQPSNSSHLDLSLAALLHGFTGGSGAP 169

Qy        167 MVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDD----------GVV 216
Db        170 LSTAALNDSTPHPRNLGSV-TNNSAGR--SDDGEESLYLGRLFGEDEEEDYEGELVGGVA 226

Qy        217 S--ETHTSTTNSSVRS-------------------------DPKFWVSVGAIAAGLAG 247
Db        227 NACEVEGLTTDEPFGSNCFANEVEIGDDEEESEIAEVLYKQDVDLGFSLDQEAI---ING 283

Qy        248 LAATGIAQALALTPEPDDPT-TTDPD------------QAANAAESATKD---------Q 285
Db        284 SYASGNSAATNVKSKPEDETKSSDPSISESSGFKDTDVNAENEASAASVDDIEKLKALEE 343

Qy        286 LTQEAFKNPENQKVNIDANGNAIP------SGE-----------------------LXD 315
Db        344 LQQDKDKNNENQLEDITNEWNGIPFTIDNETGEYIRLPLDELLNDVLKLSEFPLQDDLSN 403

Qy        316 DIVEQIAQQAKEAGEVARQQAV----------ESNAQAQQRYE---------------- 348
Db        404 DPVASTSQAAAAFNENQAQRIVSETGEDLLSGEGISSKQNRNEAKNKDNDPEKADGDSFS 463

Qy        349 ------------------DQHARRQEELQLSS---------------------------- 362
Db        464 VSDFEELQNSVGSPLFDLDEDAKKELDEMLQSTVPSYHHPHPHHGHPHAHPHSHHHASMH 523

Qy        363 -------------------------GIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTT 397
Db        524 HAHAHHAAAAAAAHQRAVQQANYGGGVGVGVG----VGVGVGSGTGSAFQR--QPA---- 573

Qy        398 TTTTHTVVQQQTGGIPQ-HKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEV 456
Db        574 -----------AGGFHHGHHQGRMPRLNRSVSMERLQD---FATYFSPIPS-------MV 612

Qy        457 GGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANS 516
Db        613 GG----VSDMSPYPHHY------PGYSYQASPSNGAP------GTPGQHGQY-------- 648

Qy        517 GGLRLGMGGLTSGGETAVSSVNAAPTPGP 545
Db        649 ------------GSGANATLQPPPPPPP 664
```

RESULT 10
S04722
puff 74E protein - fruit fly (Drosophila melanogaster)
N;Alternate names: ecdysone-induced protein E74B; ets-related protein E74B
C;Species: Drosophila melanogaster
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C;Accession: S04722; B34692
R;Janknecht, R.; Taube, W.; Luedecke, H.J.; Pongs, O.
Nucleic Acids Res. 17, 4455-4464, 1989
A;Title: Characterization of a putative transcription factor gene expressed in the 20-OH
A;Reference number: S04722; MUID:89315191; PMID:2501755
A;Accession: S04722
A;Molecule type: DNA
A;Residues: 1-883 <JAN>
A;Cross-references: EMBL:X15087; NID:g7513; PIDN:CAA33195.1; PID:g7514
A;Note: Gln-867 was also found
R;Burtis, K.C.; Thummel, C.S.; Jones, C.W.; Karim, F.D.; Hogness, D.S.
Cell 61, 85-99, 1990
A;Title: The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene
A;Reference number: A90912; MUID:90199900; PMID:2107982
A;Accession: B34692
A;Molecule type: DNA
A;Residues: 1-866,'Q',868-883 <BUR>
A;Cross-references: GB:M37083; NID:g157309; PID:g157310
C;Genetics:
A;Gene: FlyBase:Eip74EF
A;Cross-references: FlyBase:FBgn0000567
A;Map position: 3 74EF
C;Superfamily: ets DNA-binding domain homology
C;Keywords: alternative splicing; DNA binding; transcription regulation
F;789-869/Domain: ets DNA-binding domain homology <ETS>

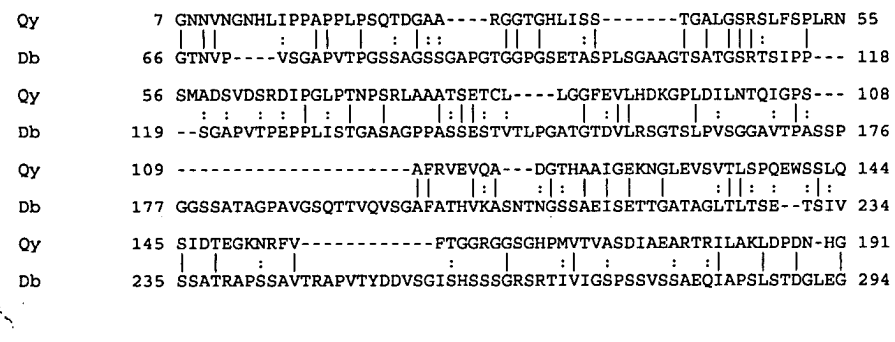
  Query Match 5.1%; Score 144; DB 2; Length 883;
  Best Local Similarity 20.8%; Pred. No. 0.35;
  Matches 124; Conservative 68; Mismatches 197; Indels 208; Gaps 26;

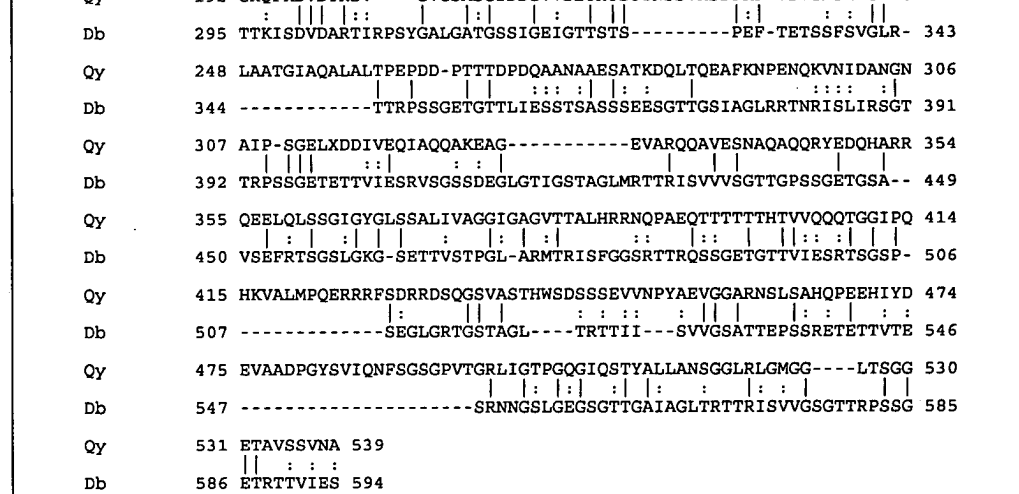
```
Qy         39 SSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHD---- 94
Db         38 SSSLSLSSSSSSSSLSSATPTPVAS---PVTPTSPPPAAAAPAEASPPAGAELQEDGQQA 94

Qy         95 ---KGPL----DILNTQIGPSAFRVEVQADG----THAAIGEKNGLEVSVTLSPQEWSSL 143
Db         95 KTQEDPTMKDQDMLEKTRQEVKDPVNVEEPGAIVDTESVMARQSPSPVASTKVPE---SL 151
```



RESULT 11
T42233
submaxillary mucin 1 - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Nov-2000
C;Accession: T42233
R;Jiang, W.; Woitach, J.T.; Keil, R.L.; Bhavanandan, V.P.
Biochem. J. 331, 193-199, 1998
A;Title: Bovine submaxillary mucin contains multiple domains and tandemly repeated, non-
A;Reference number: Z22092; MUID:98180948; PMID:9512479
A;Accession: T42233
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1589 <JIA>
A;Cross-references: EMBL:AF016589; NID:g3057086; PID:g3057087; PIDN:AAC39250.1
C;Genetics:
A;Gene: SM1
C;Superfamily: pig submaxillary mucin

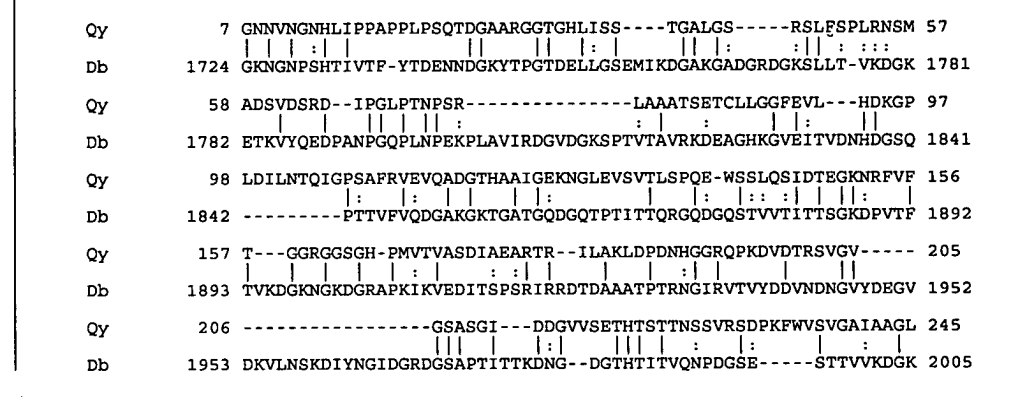
Query Match 5.1%; Score 144; DB 2; Length 1589;
Best Local Similarity 19.5%; Pred. No. 0.81;
Matches 119; Conservative 87; Mismatches 247; Indels 156; Gaps 25;





RESULT 12
B98047
hypothetical protein spr1403 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: B98047
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: B98047
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2551 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00207.1; PID:g15459054; GSPDB:GN00174
C;Genetics:
A;Gene: spr1403

Query Match 5.1%; Score 143.5; DB 2; Length 2551;
Best Local Similarity 21.3%; Pred. No. 1.7;
Matches 131; Conservative 75; Mismatches 245; Indels 165; Gaps 32;



```
Qy       246 AGLAATGIAQALALTPEPDDP---TTTDPDQAANAAESATKDQLTQEAFKNPENQKVNID 302
Db      2006 DGKTAN-----ITTTENPDGSHTITVTNPD-------GSTKETVVKNG-KDGKTPKVEVT 2052

Qy       303 ANGNAIPSGELXD---DIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQ 359
Db      2053 DNNDGTHTVKVTDGDGNVTNAIIKDGKD-GKAATATTTENPDGSHTVTITNPDGTKNEFV 2111

Qy       360 LSSGIGYGLSSALIVAGGIGAGVTTALHRRN----------QPAEQTTTTTTH------- 402
Db      2112 VKNG-----------RDGVDGRTPTASVRDNGDGSHTIVITNPEGVTTETTVRDGKSPKV 2160

Qy       403 TVVQQQTGGIPQHKVALMPQE-----------RRRFSDRRDSQ-----------GSVAS 439
Db      2161 TITDEQNG---THKISVLNGDGTTTETIIKDGKSPVATVRDNQDGTYTIRVENGNGTVSE 2217

Qy       440 THWSDSSSEVVNPYAEV---GGARNSLSAHQPEEHIYDEVAAD---PGYSVIQNFSGSGP 493
Db      2218 TTVRDGKS----PTAKVVDNGDGTHTITVVNSDGITTTTTVRDGREPKLEVIDNNDGSHT 2273

Qy       494 VTGRLIGTPGQGIQST 509
Db      2274 I--KVTGADGKGTTTT 2287
```

RESULT 13
T39141
hypothetical protein SPAC8C9.04 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39141
R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21748
A;Accession: T39141
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-647 <OLI>
A;Cross-references: EMBL:Z99168; PIDN:CAB16292.1; GSPDB:GN00066; SPDB:SPAC8C9.04
A;Experimental source: strain 972h-; cosmid c8C9
C;Genetics:
A;Gene: SPDB:SPAC8C9.04
A;Map position: 1

  Query Match   5.1%; Score 141.5; DB 2; Length 647;
  Best Local Similarity   19.0%; Pred. No. 0.32;
  Matches 103; Conservative   64; Mismatches 207; Indels 169; Gaps   20;

```
Qy       114 VQADGTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASD 173
Db        94 LQEPGFTQTIVEKDADQVDEPLEPIASSALGTVEPPTDNK-----------PSASTSTA 141

Qy       174 IAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTN-------- 225
Db       142 VPTTEARNTSITEPANSPSSSSSSASTKSTATTQSA---DYVVAEHFAPQRNDEQLGNSP 198

Qy       226 SSVRSDPKFWVSV---GAIAAGLAGLAATGIAQALALTPE--PDDP--------TTTDPD 272
Db       199 ASITSKPATTSAAQPSSKVEENMAKATSQPITTAEKEIPELKPIEPEAIMISKEINTTHD 258

Qy       273 QAA--------------------------NAAESATKDQLTQEAFKN----------P 294
Db       259 QAAATTTAAVASASTTATAESHAVADGIMNDNVLESIGENVQQETVFEDASDIPHADVIP 318

Qy       295 ENQKVNID--------------------ANGNAIPSG--ELXDDIVEQIAQQAKEAGEV 331
Db       319 HTTTVTVEEESPIALGQGGVTHEATTSARASASGIPGAFEEVQQTVQEDLPHPTAETVEI 378

Qy       332 ARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSAL----------------IV 374
Db       379 ARFAEQPVRAQQPEQYESSVV--QEATETVTDVGKGVSSTVKNEVNVPSTIPTESENPVA 436

Qy       375 AGGIGAGVTTALHRRNQ----PAE------QTTTTTTHTVVQQQTGGIPQHKVALMPQER 424
Db       437 VGG-----TTAEHPVQEAVTAPTETAHDFSKETTTASKRVSKHDKASAEKHKVARKP--- 488

Qy       425 RRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSV 484
Db       489 ------------------SSTGQEPTTPSTPAKSAQSSKHARRPSKQ-----ASAP---- 521

Qy       485 IQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAA---P 541
Db       522 ----SSPGTTSAAVPGGKKSAIEAA-APIPTSADTVESKHAAGSGSATTIPSPGSATTKP 576

Qy       542 TPG 544
Db       577 TPG 579
```

RESULT 14
AC0304
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0304
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2535 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2490

  Query Match   5.0%; Score 141; DB 2; Length 2535;
  Best Local Similarity   21.9%; Pred. No. 2.4;
  Matches 129; Conservative   73; Mismatches 246; Indels 140; Gaps   25;

```
Qy        25 QTDGA---ARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATS 81
Db      1485 QTDGALTLAAGGDVLLTTQTEQHDEQRNHTGLSKGIASSTLTRTEDSL----------S 1533

Qy        82 ETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQAD----------GT--------HAA 122
Db      1534 QTLAVGS---MLSAGSIDVSGKNIAVMGSNVVADQDISLRAQENITVGTAQQSESESHLF 1590

Qy       123 IGEKNGL----EVSVTL---SPQEWSSLQSIDTEGKNRFVFTGGRGGS--GHPMVTVASD 173
Db      1591 EQKKSGLMSTGGIGVTVGSSSTKMTDSGQSISS--------VGSTVGSVLGNVSMTAGED 1642

Qy       174 IAEARTRILAKLDPDNHGGR----------QPKDVDTRSVGVGSA-SGIDDGVVSETHT 221
Db      1643 LRVQGAEVLAGKDINLTGKNVSILAAENQLTQSHTVEQKQSGLTLALSGAVGSAVNTAVT 1702

Qy       222 STTNSSVRSDPKFWVSVGAIAA--GLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAE 279
Db      1703 TAKAASEESSGRLGALQGVKAALNGVQAVQAGQLVQA-----EGGDAASMFGISASLGSQ 1757

Qy       280 SATKDQ------LTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEV-- 331
Db      1758 KSSSEQHQEQTHVTGSTLTAGNNLTINATGEGNAANSG----DIVVQ-GSQLQAGGDTTL 1812

Qy       332 -ARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRN 390
Db      1813 DAARDVLLLGAANTQKTDGSNSSSGGSVGVSLGIS-GASSGLSIFANANKG------QGS 1865

Qy       391 QPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSS---S 447
Db      1866 EHGDGTSWTET----------------------------TLDSGGTLSLYSGRDTSLVGA 1897
```

```
Qy       448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNPSGSGPVTGRLIGTPGQ--- 504
Db      1898 QVSGETVKVEVGRDLLLQSQQDSDNYD--AKQQNSSVGGSFS-PGSMTGSISINGSQDKL 1954

Qy       505 -----GIQSTYALLANSGGLRLGMGGLTS-----GGETAVSSVNAAPT 542
Db      1955 HSNFDSVQEQTGIFAGSGGFDITVGGHTQLDGAVIGSTATADKNTLDT 2002
```

RESULT 15
AE1852
hypothetical protein alr0366 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AE1852
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE1852
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1152 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB72324.1; PID:g17129711; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr0366

Query Match 5.0%; Score 140; DB 2; Length 1152;
Best Local Similarity 21.2%; Pred. No. 0.91;
Matches 127; Conservative 90; Mismatches 236; Indels 146; Gaps 29;

```
Qy         5 NLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRS----LFSPLRNSMADS 60
Db       124 NASLNIGGSFVATTANTI--QFPGGAEFSLTSPVTSTNTLLSVNPTAFLFNQIANQGTNS 181

Qy        61 VDSRDIPGLPTNPSRLAAATSETCLLGG---------FEVLHDKGPLDILNTQIGPSAFR 111
Db       182 IENRGYLAVPNNKSLI--------LLGGNIAPTSNATGKILIDGGVVQALNG-------R 226

Qy       112 VEVQADGTHAAIG-EKNGLEVSVTLSPQEWSSLQSIDTEG-KNRFVFTGGRGGSGHPMVT 169
Db       227 VEIGGLVEPGFIGINVDGNQLSLTFP----DSVAKTDISSINNGTVFTSGAGGG--DIVV 280

Qy       170 VASDIAEAR-----TRILAKLDPDNHGG--RQPKDVDTRSVGV----------GSASGID 212
Db       281 NADNLSLLNYGAFFTGIL-----NNQGNAETQAGDISINATGIVTVAQNSIISNSSLGIG 335

Qy       213 D----GVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTT 268
Db       336 DSGKINIVAQSLQIIDNSSVQS------------FSLQGNSGT-------VNVKVDDTVS 376

Qy       269 TDPDQAANAA--ESATKDQLTQEAFKNPENQK---VNIDA------NGNAIPSGELXDDI 317
Db       377 LLSGQISSSVRPRGASASSLQSNLLGTPTRGKSGGINIQARSLLAVDSSAISASNFLADD 436

Qy       318 VEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGG 377
Db       437 SGDIKIQATDAVILNNRSTISSSAFGQ--------GKSGNLSINTNWLNIINNSQITANT 488

Qy       378 IGAG-------VTTALHRRNQPAEQTTTTTTHTVVQQQTGG---IPQHKVALMPQERRRF 427
Db       489 LGTGNAGDINIIALDINIDKSLITGSTSSFLNTIVNLGNAGNINIQTARINL-------- 540

Qy       428 SDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAAD--PGYSVI 485
Db       541 ----TNSGFIISTSGRPEQKET----SGFGGNINITATELIE--IDPKGATDIITGFST- 589

Qy       486 QNFSGS--GPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAAPT 542
Db       590 RTFSGSRAGDIT-----------LNTKNLIVRNGGVIMADAANNLGGNAGNININASDT 637
```

Search completed: December 16, 2003, 08:56:49
Job time : 30.2608 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:55:50 ; Search time 26.7563 Seconds
(without alignments)
3816.109 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

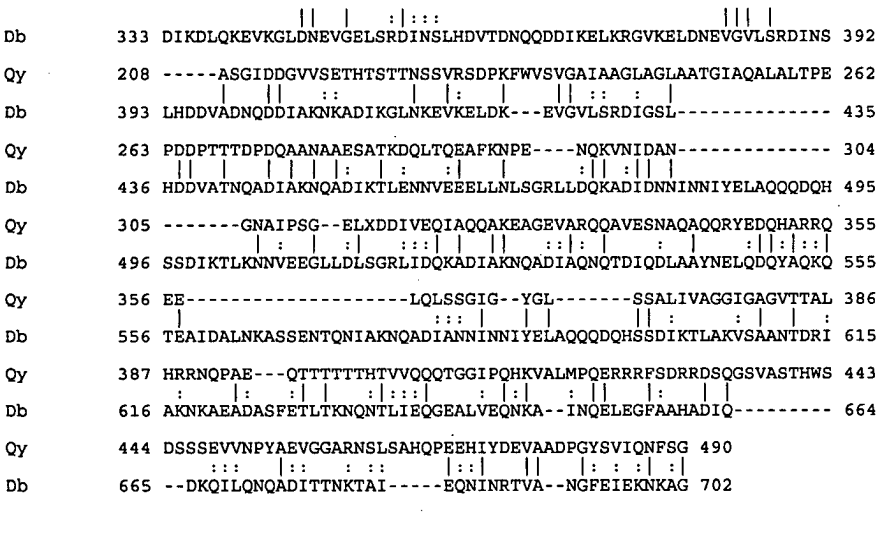
Result No.	Score	% Query Match	Length	DB	ID	Description
1	149	5.3	831	11	US-09-952-267-1	Sequence 1, Appli
2	141.5	5.1	1488	10	US-09-738-626-5495	Sequence 5495, Ap
3	139	5.0	966	12	US-10-304-454-2	Sequence 2, Appli
4	138.5	4.9	2353	10	US-09-797-862-33	Sequence 33, Appl
5	137.5	4.9	1829	15	US-10-156-761-10049	Sequence 10049, A
6	137	4.9	2468	12	US-10-246-330-4	Sequence 4, Appli
7	136	4.9	1569	12	US-10-287-274-312	Sequence 312, App
8	136	4.9	2086	9	US-09-815-242-5639	Sequence 5639, Ap
9	136	4.9	5795	9	US-09-815-242-12610	Sequence 12610, A
10	136	4.9	7746	15	US-10-156-761-7965	Sequence 7965, Ap
11	135.5	4.8	1138	15	US-10-074-475-194	Sequence 194, App
12	135	4.8	2344	9	US-09-815-242-12713	Sequence 12713, A
13	133	4.8	864	15	US-10-156-761-12878	Sequence 12878, A
14	132.5	4.7	1026	9	US-09-379-931-7	Sequence 7, Appli
15	132.5	4.7	1026	12	US-10-223-597-7	Sequence 7, Appli
16	132.5	4.7	5215	10	US-09-861-289-2	Sequence 2, Appli
17	132.5	4.7	5215	10	US-09-860-846-2	Sequence 2, Appli
18	132.5	4.7	5215	11	US-09-988-384B-2	Sequence 2, Appli
19	132.5	4.7	5215	11	US-09-836-821-2	Sequence 2, Appli
20	132.5	4.7	5215	12	US-10-271-889-45	Sequence 45, Appl
21	132.5	4.7	6146	15	US-10-156-761-10436	Sequence 10436, A
22	131.5	4.7	707	15	US-10-128-714-8571	Sequence 8571, Ap
23	130	4.6	1260	15	US-10-245-802-8	Sequence 8, Appli
24	129.5	4.6	2834	12	US-10-085-959-252	Sequence 252, App
25	127.5	4.6	771	12	US-10-205-219-175	Sequence 175, App
26	126.5	4.5	1089	12	US-10-379-632-109	Sequence 109, App
27	126.5	4.5	1209	12	US-10-379-632-96	Sequence 96, Appl
28	126.5	4.5	1684	12	US-10-238-075-877	Sequence 877, App
29	126	4.5	376	12	US-10-084-843-202	Sequence 202, App
30	126	4.5	376	12	US-10-193-002-197	Sequence 197, App
31	126	4.5	1073	15	US-10-156-761-12156	Sequence 12156, A
32	125.5	4.5	1108	12	US-09-949-029-108	Sequence 108, App
33	125	4.5	2442	14	US-10-109-886-10	Sequence 10, Appl
34	124.5	4.4	729	9	US-09-287-849-2	Sequence 2, Appli
35	124.5	4.4	729	12	US-10-359-460-2	Sequence 2, Appli
36	124.5	4.4	729	12	US-10-098-732A-16	Sequence 16, Appl
37	124.5	4.4	2514	12	US-10-320-800-40	Sequence 40, Appl
38	124	4.4	506	15	US-10-156-761-12648	Sequence 12648, A
39	124	4.4	1180	12	US-10-193-764-61	Sequence 61, Appl
40	124	4.4	1188	12	US-10-193-764-59	Sequence 59, Appl
41	124	4.4	1536	12	US-10-193-764-63	Sequence 63, Appl
42	123.5	4.4	1270	15	US-10-245-802-16	Sequence 16, Appl
43	122.5	4.4	507	9	US-09-910-087-19	Sequence 19, Appl
44	122.5	4.4	729	12	US-10-098-732A-18	Sequence 18, Appl
45	122.5	4.4	930	12	US-10-098-732A-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-09-952-267-1
; Sequence 1, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-1

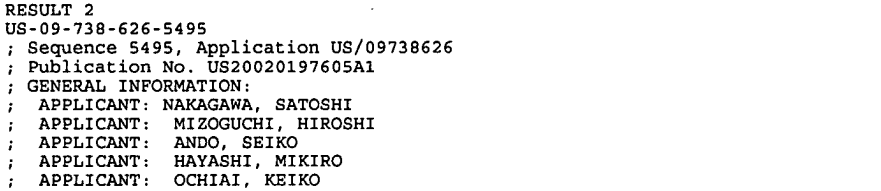
Query Match 5.3%; Score 149; DB 11; Length 831;
Best Local Similarity 19.2%; Pred. No. 0.0054;
Matches 101; Conservative 73; Mismatches 193; Indels 160; Gaps 21;

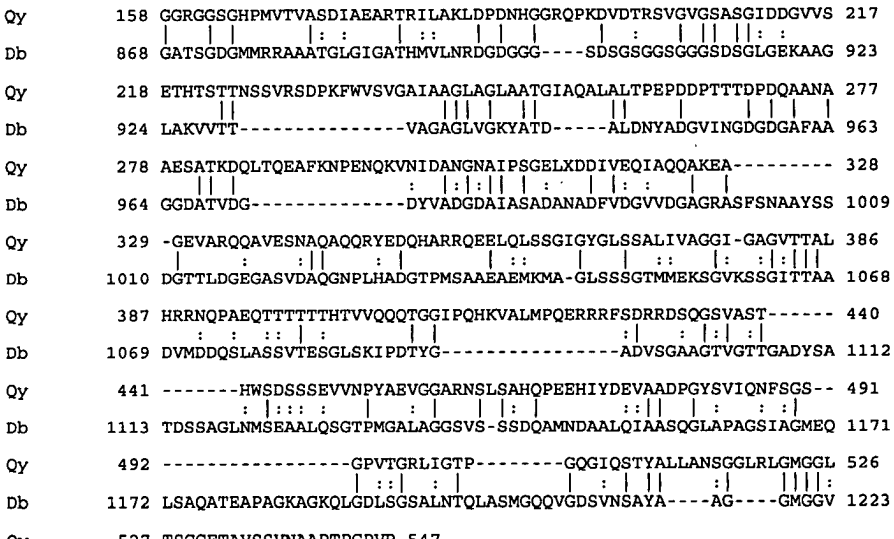
```
Qy        74 SRLAAATSETCLLGGFEVLHDKGPLDILNTQIGP---SAFRVEVQADGTHAAIGEKNGLE 130
             |     || : |||       |    : | ::|    : |  | :|:    ::| : |
Db       226 SGTTGVTSNSVLLGNETA--GKQATTVKNAEVGGLSLTGFAGESKAENGVVSVGSEGGER 283

Qy       131 VSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVAS----DIAEARTRILAKLD 186
               | :     :   |:  |       |  | : ||      ||  |  ||   |
Db       284 QIVNVGA---GQISDTSTDAVN--------GSQLHALATVVDDNQYDIVNNRADILNNQD 332

Qy       187 P----------DNHGGRQPKDVDT----------------------RSVGVGS----- 207
```



RESULT 2
US-09-738-626-5495
; Sequence 5495, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5495
; LENGTH: 1488
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5495

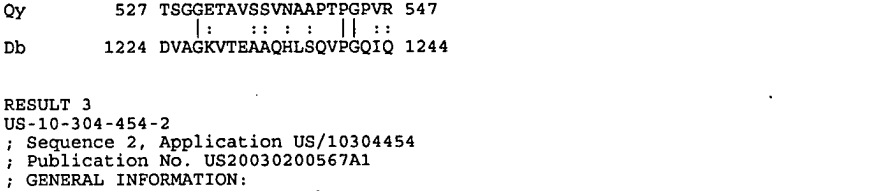
Query Match 5.1%; Score 141.5; DB 10; Length 1488;
Best Local Similarity 18.9%; Pred. No. 0.049;
Matches 106; Conservative 75; Mismatches 225; Indels 155; Gaps 19;





RESULT 3
US-10-304-454-2
; Sequence 2, Application US/10304454
; Publication No. US20030200567A1
; GENERAL INFORMATION:
; APPLICANT: Komatsu, Setsuko
; APPLICANT: Sharma, Arun
; APPLICANT: Hashimoto, Junji
; APPLICANT: Sakaguchi, Kengo
; TITLE OF INVENTION: Cold Stress-Responsive CRTintP Gene And
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 3462.1001-000
; CURRENT APPLICATION NUMBER: US/10/304,454
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: JP 2002-121275
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 966
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-304-454-2

Query Match 5.0%; Score 139; DB 12; Length 966;
Best Local Similarity 20.9%; Pred. No. 0.041;
Matches 128; Conservative 82; Mismatches 223; Indels 178; Gaps 31;



```
Qy        172 ----SDIAEARTRILAKLDPDNHG-----GRQPKDVDTRSVGVGSASGID--DGVVSETH 220
Db        498 ASIPSIVAQVTAQVANALSANQQGQVSSSAQNTVDQGSRSV---TTNGVDNVDSLVS--- 551

Qy        221 TSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAES 280
Db        552 ASTQLQNELSDSN---------NGRTSLNAQSLVAGAGISPS----NTSDPN-------L 591

Qy        281 ATKDQLTQEA------FKNPENQKVNID---------ANGNAIPSGELXDDIVEQIAQQA 325
Db        592 ASEDSSTENAPNIGSIQQHPEMEGIHADNVRKPSEESTTANLV--GQITTTCTDDIS--V 647

Qy        326 KEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGI------------------GYG 367
Db        648 NRSAENSSQKNIPLDGVSAQSIKPSASSRSEPVGLGGGLQPKRRSRTAKPPGSSSDTGEV 707

Qy        368 LSSALIV----AGGIGAGVTTALHRRNQPAEQT--TTTTTHTVVQQQTGGIPQHKVALMP 421
Db        708 VNSSRISNSQNAVSMGQQVLQALASQNTNVNRSHVTDSPLPSTTSQFSGGMPPRR----- 762

Qy        422 QERRRFSDRRDSQGSVASTHWSDSSSEVVNPYA---------------EVGGARNSLSAH 466
Db        763 ------------QGGEGQVDFGSMISSVLNNPAFGNLLSNVAEQTGMGSAGDLRNMV--- 807

Qy        467 QPEEHIYDEVAADPGYSVIQNFSGSG----------------PVTGRLIGTPGQGIQSTY 510
Db        808 --EECAQSPAIMDTMSNLVQNVDGSGRGQGGIDLSRMMQQMMPVVSQVLGGAGARPAGT- 864

Qy        511 ALLANSGGLRL 521
Db        865 ----NSGQSRL 871
```

RESULT 4
US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

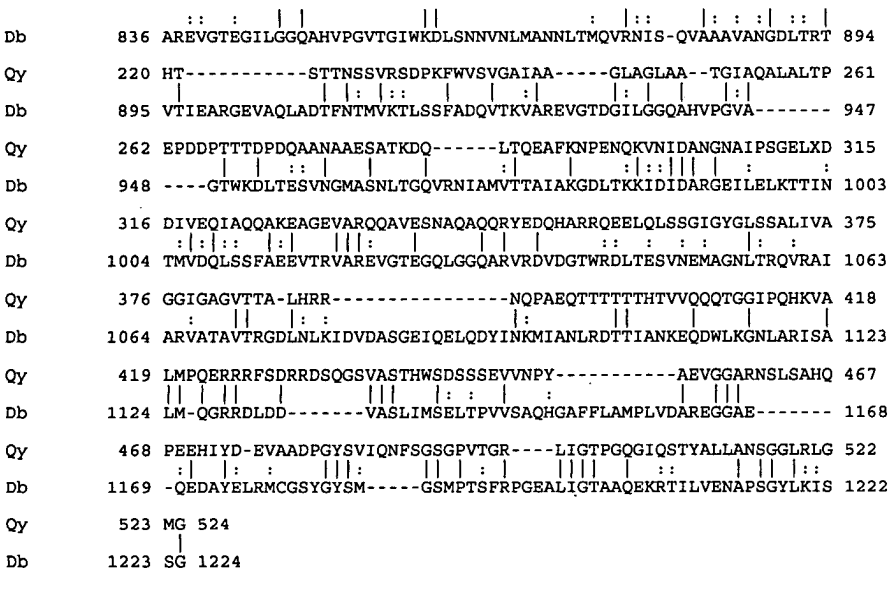
Query Match 4.9%; Score 138.5; DB 10; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.17;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

```
Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
Db       1163 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1198

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
Db       1199 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1243

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
Db       1244 AV--KNANEVEFV--------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1284

Qy        179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
Db       1285 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1343

Qy        230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
Db       1344 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1396

Qy        278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
Db       1397 ANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1444

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
Db       1445 LGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
Db       1498 ----------------------GG-----NGITDNEKKRAASVKD--------------- 1515

Qy        448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
Db       1516 -VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1569

Qy        508 STYALLANSGGLRLGMGGLTSGGE 531
Db       1570 TEVKIGAKTSVIKDHNGKLFTGKE 1593
```

RESULT 5
US-10-156-761-10049
; Sequence 10049, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10049
; LENGTH: 1829
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10049

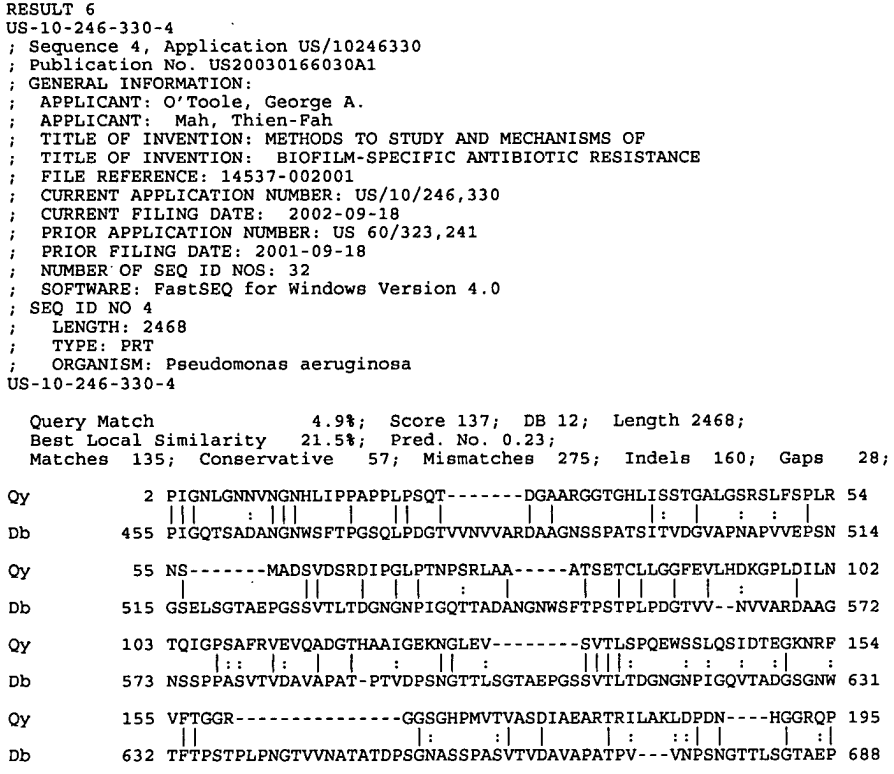
Query Match 4.9%; Score 137.5; DB 15; Length 1829;
Best Local Similarity 20.9%; Pred. No. 0.14;
Matches 126; Conservative 84; Mismatches 257; Indels 135; Gaps 23;

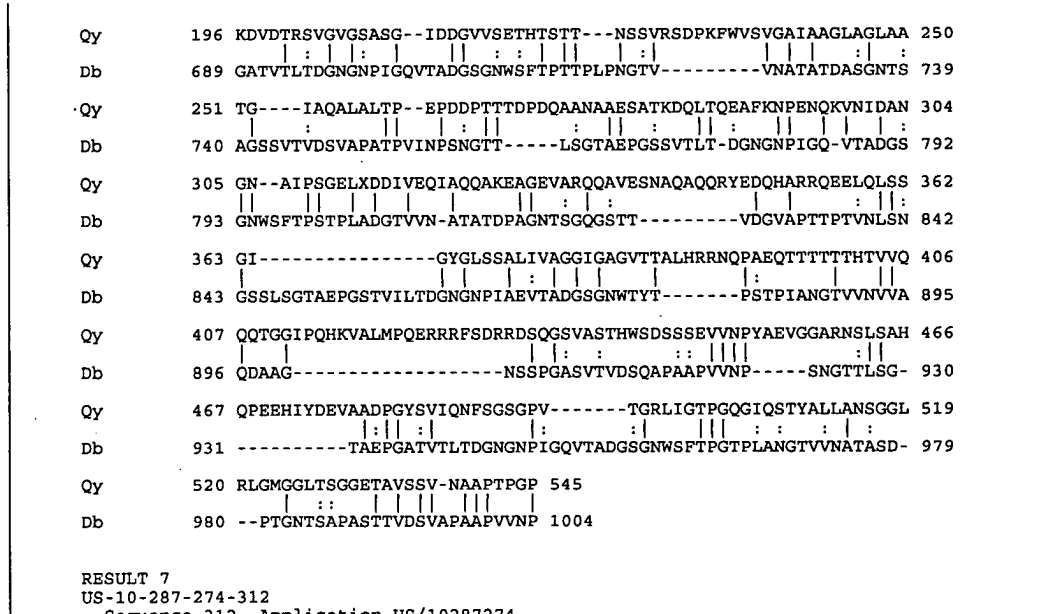
```
Qy          5 NLGNNVNG--NHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVD 62
Db        676 DLTDSVNGMAGNLTAQVRNIAQVATAVARGDLSQKID----VDARGEILELKNTLNTMVD 731

Qy         63 SRDIPGLPTNPSRLAAATSETCLLGG-FEVLHDKGPLDILNTQIGPSAFRVEVQADGTHA 121
Db        732 --QLSNFAEQVTRVAREVGTEGMLGGQAEVQGVSGTWKDLTQSVNFMANNLTIQV----- 784

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTRI 181
Db        785 ----RNIAEVTTAVAKGDLSKKITVDAKGEILELVT-----TVNTMVDQLSSFAEQVTRV 835

Qy        182 LAKLDPDNHGGRQP---------KD-------------VDTRSVGVGSASGIDDGVVSET 219
```



RESULT 6
US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

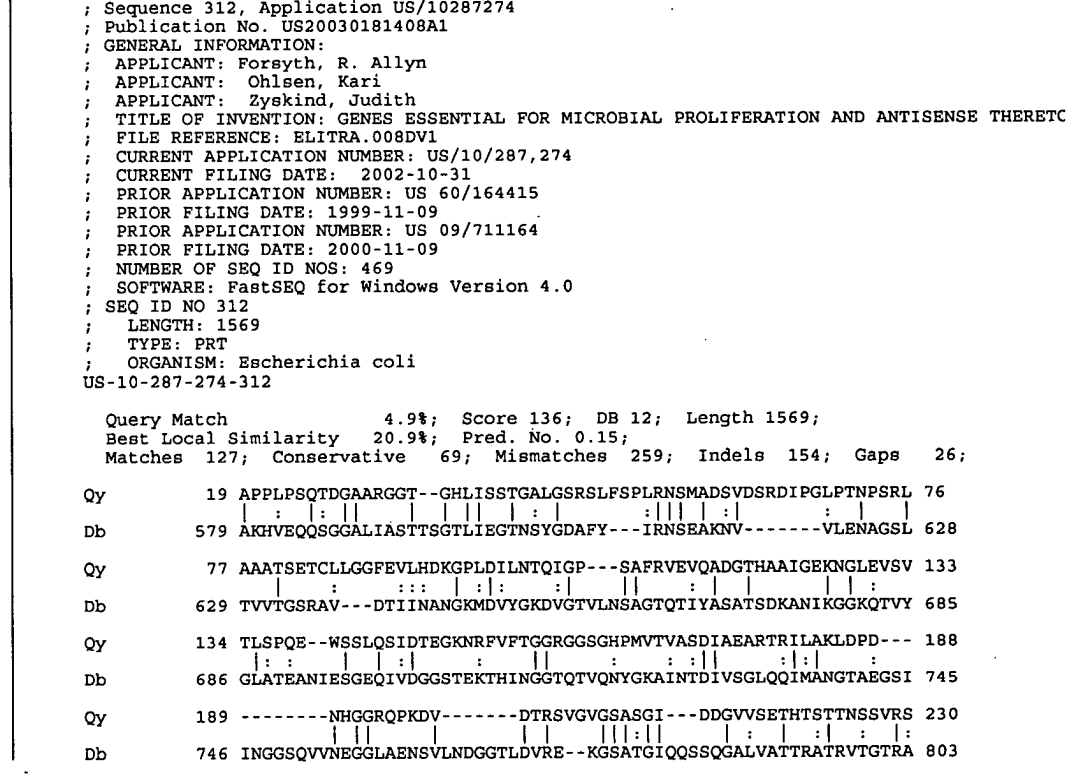
Query Match 4.9%; Score 137; DB 12; Length 2468;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 135; Conservative 57; Mismatches 275; Indels 160; Gaps 28;





RESULT 7
US-10-287-274-312
; Sequence 312, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETC
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 312
; LENGTH: 1569
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-287-274-312

Query Match 4.9%; Score 136; DB 12; Length 1569;
Best Local Similarity 20.9%; Pred. No. 0.15;
Matches 127; Conservative 69; Mismatches 259; Indels 154; Gaps 26;



```
Qy        231 DPKFWVSVGAIAAGLAGLAATGI--AQALALTPEPDDPTTTDPDQAANAAESATKDQLTQ 288
Db        804 D--------GVAFSIEQGAANNILLANGGVLTVESD--TSSDKTQVNMGGREIVKTKAT- 852

Qy        289 EAFKNPENQKVNIDANGNAIPSGELXDDIVEQIA----------QQAKEAGEVARQQ--- 335
Db        853 --------------ATGTTLTGGE---QIVEGVANETTINDGGIQTVSANGEAIKTKINE 895

Qy        336 ----AVESNAQAQQRYEDQHARRQEE----LQLSSGIGYGLSS----------------A 371
Db        896 GGTLTVNDNGKATDIVQNSGAALQTSTANGIEISGTHQYGTFSISGNLATNMLLENGGNL 955

Qy        372 LIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRR 431
Db        956 LVLAGTEARDSTVGKGGAMQNLGQDSATKVN------SGG--QYTLGRSKDEFQALARAE 1007

Qy        432 DSQ--GSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFS 489
Db       1008 DLQVAGGTAIVYAGTLAD------ASVSGATGSLSLMTPRDNV----------TPVKLE 1050

Qy        490 GSGPVTGRLIGTPGQGIQSTYALL--ANSGGLRLGMGGLTSG--------------GET 532
Db       1051 GAVRITDSATLTLGNGVDTTLADLTAASRGSVWLNSNNSCAGTSNCEYRVNSLLLNDGDV 1110

Qy        533 AVSSVNAAP 541
Db       1111 YLSAQTAAP 1119
```

RESULT 8
US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5639

Query Match 4.9%; Score 136; DB 9; Length 2086;
Best Local Similarity 19.2%; Pred. No. 0.22;
Matches 113; Conservative 89; Mismatches 243; Indels 144; Gaps 25;

```
Qy         53 LRNSMADSVDSRDIPGLPT----NPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPS 108
Db       1381 LRNTKTSATNT--INGLPNLTQLQKDNLKHQVEQAQNVAGVNGVKDKG--NTLNTAMG-- 1434

Qy        109 AFRVEVQADGT-----------------HAAIGEKNGLEVSVTLSPQEWSSLQSIDTEG 150
Db       1435 ALRTSIQNDNTTKTSQNYLDASDSNKNNYNTAVNNANGV-INVTNNP-------NMDANA 1486

Qy        151 KNRFVFTGGRGGSGHPMVTV------ASDIAEARTRILAKLDPDNHGGRQPKDVDTRSV- 203
Db       1487 IN---------GMANQVNTTKAALNGAQNLAQAKTTAKNNLNNLTSINNAQKDALTRSID 1537

Qy        204 GVGSASGID-----------------DGVVSETHTSTTNSSVRSDPK----FWVSVGAIA 242
Db       1538 GATTVAGVNQETAKATELNNAMHSLQNGINDETQTKQTQKYLDAGPSKKSAYDQAVNAAK 1597

Qy        243 AGLAGLAATGIAQAL---ALTPEPDDPTTTDPDQAANAAESATKDQL---------TQEA 290
Db       1598 AILTKASGQNVDKAAVEQALQNVNSTKTALNGDAKLNEAKAAAKQTLGTLTHINNAQRTA 1657

Qy        291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIA---------------QQAKEAGEVARQQ 335
Db       1658 LDNEITQATNVEGVNTVKAKAQQLDGAMGQLETSIRDKDTTLQSQNYQDADDAKRTAYSQ 1717

Qy        336 AVESNAQAQQR-------YEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTA--L 386
Db       1718 AVNAAATILNKTAGGNTPKADVERAMQAVTQANTALN-GIQNLERAKQAANTAITNASDL 1776

Qy        387 HRRNQPAEQTTTTTT----------HTVVQQQTGGIP-QHKVALMPQ------------ 422
Db       1777 NTKQKEALKAQVTSAGRVSAANGVEHTATELNTAMTALKHAIADKAETKASGNYVNADAN 1836

Qy        423 ERRRFSDRRDSQGSVASTHWSD--SSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP 480
Db       1837 KRQAYDEKVTAAENIVSGTPTPTLTPSDVTNAATQVTNAKTQLNGNHNL-----EVAKQN 1891

Qy        481 GYSVIQNFSG-SGPVTGRLIGTPGQG--IQSTYALLANSGGLRLGMGGL 526
Db       1892 ANTAIDGLTSLNGPQKAKLKEQVGQATTLPNVQTVRDNAQTLNTAMKGL 1940
```

RESULT 9
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308

```
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12610
;    LENGTH: 5795
;    TYPE: PRT
;    ORGANISM: Staphylococcus aureus
US-09-815-242-12610

  Query Match             4.9%;  Score 136;  DB 9;  Length 5795;
  Best Local Similarity   19.2%;  Pred. No. 0.98;
  Matches  113;  Conservative   89;  Mismatches  243;  Indels  144;  Gaps   25;

Qy          53 LRNSMADSVDSRDIPGLPT----NPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPS 108
               |||:   : ::  | |||         |     :   : |   : |||  : ||| :|
Db        3658 LRNTKTSATNT--INGLPNLTQLQKDNLKHQVEQAQNVAGVNGVKDKG--NTLNTAMG-- 3711

Qy         109 AFRVEVQADGT-----------------HAAIGEKNGLEVSVTLSPQEWSSLQSIDTEG 150
               | |  :| | |                 : |:   ||: ::|| :|       ::|
Db        3712 ALRTSIQNDNTTKTSQNYLDASDSNKNNYNTAVNNANGV-INVTNNP-------NMDANA 3763

Qy         151 KNRFVFTGGRGGSGHPMVTV------ASDIAEARTRILAKLDPDNHGGRQPKDVDTRSV- 203
                 |          |  : : |       | ::|:|:|     |:          ||  |||:
Db        3764 IN---------GMANQVNTTKAALNGAQNLAQAKTTAKNNLNNLTSINNAQKDALTRSID 3814

Qy         204 GVGSASGID----------------DGVVSETHTSTTNSSVRSDPK----FWVSVGAIA 242
               |  : :|::                    :|:  || |  |   : : |      :  :| |
Db        3815 GATTVAGVNQETAKATELNNAMHSLQNGINDETQTKQTQKYLDAGPSKKSAYDQAVNAAK 3874

Qy         243 AGLAGLAATGIAQAL---ALTPEPDDPTTTDPDQAANAAESATKDQL---------TQEA 290
               | |   :  : :|    ||       | : |   | |::| | |           : |
Db        3875 AILTKASGQNVDKAAVEQALQNVNSTKTALNGDAKLNEAKAAAKQTLGTLTHINNAQRTA 3934

Qy         291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIA---------------QQAKEAGEVARQQ 335
                 |   |  |::          :  |  : |:                 | | :|   |  |
Db        3935 LDNEITQATNVEGVNTVKAKAQQLDGAMGQLETSIRDKDTTLQSQNYQDADDAKRTAYSQ 3994

Qy         336 AVESNAQAQQR-------YEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTA--L 386
               || : |    :          |      |   | :: :  |: :           :| |  |
Db        3995 AVNAAATILNKTAGGNTPKADVERAMQAVTQANTALN-GIQNLERAKQAANTAITNASDL 4053

Qy         387 HRRNQPAEQTTTTTT----------HTVVQQQTGGIP-QHKVALMPQ------------- 422
               : : : | :   |:           ||  :  |     :| :|  :
Db        4054 NTKQKEALKAQVTSAGRVSAANGVEHTATELNTAMTALKHAIADKAETKASGNYVNADAN 4113

Qy         423 ERRRFSDRRDSQGSVASTHWSD--SSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP 480
               :|: : ::  :  :: |   :  : |:| |   :|  |:  |: :          |||
Db        4114 KRQAYDEKVTAAENIVSGTPTPTLTPSDVTNAATQVTNAKTQLNGNHNL-----EVAKQN 4168

Qy         481 GYSVIQNFSG-SGPVTGRLIGTPGQG--IQSTYALLANSGGLRLGMGGL 526
                 : |   :  :||   :|    ||   : :   :   |:  |   | ||
Db        4169 ANTAIDGLTSLNGPQKAKLKEQVGQATTLPNVQTVRDNAQTLNTAMKGL 4217


RESULT 10
US-10-156-761-7965
; Sequence 7965, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7965
;    LENGTH: 7746
;    TYPE: PRT
;    ORGANISM: Streptomyces avermitilis
US-10-156-761-7965

  Query Match             4.9%;  Score 136;  DB 15;  Length 7746;
  Best Local Similarity   19.3%;  Pred. No. 1.5;
  Matches  116;  Conservative   79;  Mismatches  241;  Indels  166;  Gaps   24;

Qy           1 MPIGNLGNNVNGNH---LIPPAPPLPSQTDGAARGGTG-HLISSTGAL----------G 45
               :|:  |   |   |         || :      |||  |:: ::||           |
Db         401 IPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVEAGEGATAGGG 460

Qy          46 SRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQI 105
               :   | ||         ::   ||    : | |               | : || :    :
Db         461 AAPTFWPL--------SAKTASGLRAQAAALLA-------------HTEAPLGLALPDV 498

Qy         106 GPSAFRVEVQADGTHAAIGE-------------KNGLEVSVTLSPQEWSSLQSIDTEGKN 152
               | |         :    :||              |:: ::|               |:
Db         499 GWSLATGRTAFEHRAVVVGEVRDDFLRALRELSTGGIDAALTTG--------RAGPRGEL 550

Qy         153 RFVPTG---GRGGSGHPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSAS 209
               |:|:|   | | |  :      | |   : | |||      |  ::|    || | |:
Db         551 AFLFSGQGSQRAGMGRELAESVPAFASALDEVCAHLDP--LLPRPLREVMFAPVGTGEAA 608

Qy         210 GIDDGVVSETHTSTTNSS---------VRSDPKFWVSVGAIAAG-LAGLAATGIAQALAL 259
               :|  : ::|       :          :  |      |:| :||  :||: :  :| | ||
Db         609 ELDQTLYTQTSLFAVEVALFRLLESWDITPDVLMGHSIGELAAAHVAGVLS--LADACAL 666

Qy         260 TPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDA-NG--NAIPSG--ELX 314
                         |   |  | | :|::   |        :|:| | ||  : : ||  :|
Db         667 VAARGRLMQALPAGGAMIAVQAAEDEV--RARIGERTDRVSIAALNGPDSVVVSGDEDLA 724

Qy         315 DDIVEQIAQQAKEAGEVARQQAVES---------------------------------- 339
               :| :  |   ::   :   |  |
Db         725 TEIADAFAALGRKTSRLRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLS 784

Qy         340 ----------------NAQAQQRYEDQHARRQEE-----LQLSSGIGYGLSSALIVAGGI 378
                             : :   |: |  ||  |:     |:|  |   |: :::  ||
Db         785 EEPSGYEGCTAAYWVRHVREAVRFADGVARLDEQGVRTYLELGPG---GVLASMARAGAD 841

Qy         379 GAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGI-------------PQHKVALMPQ--E 423
                  :    |  :|  |          |     |:              : ||||     |
Db         842 SESLFVPALRTRRPEAQALLA---AVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFE 898

Qy         424 RRRFSDRRDSQGSVASTHWS------DSSSEVVN-PYAEVGGARNSLSAHQPEEHIYDEV 476
               |:|:    |:   ||:| |:      |: :|  : |  :|     |:  : :| : | |
Db         899 RQRYWLDGDTTSPVATTGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMV 958

Qy         477 AA 478
               ||
Db         959 RA 960


RESULT 11
US-10-074-475-194
; Sequence 194, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT:  Macina, Roberto
;  APPLICANT:  Hu, Ping
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Karra, Kalpana
;  APPLICANT:  Cafferkey, Robert
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
;  TITLE OF INVENTION:  Genes and Proteins
;  FILE REFERENCE: DEX-0313
```

```
;  CURRENT APPLICATION NUMBER: US/10/074,475
;  CURRENT FILING DATE:  2002-02-13
;  PRIOR APPLICATION NUMBER: 60/268,292
;  PRIOR FILING DATE: 2001-02-13
;  NUMBER OF SEQ ID NOS: 295
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 194
;   LENGTH: 1138
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-074-475-194

  Query Match             4.8%;  Score 135.5;  DB 15;  Length 1138;
  Best Local Similarity   21.8%;  Pred. No. 0.1;
  Matches   85;  Conservative   56;  Mismatches  134;  Indels  115;  Gaps   18;

Qy          8 NNVNGNHLIPPA---------PPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMA 58
              ||:  |  :| |         | || |  |    :|  | |    |  |: |  :
Db        204 NNLPSNMPLPSASPLHQLDLKPSLPLQNSGTHTPGLLEDLSKNGRLPEIKL--PV-NGCS 260

Qy         59 DSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKG----PLDILNTQIGPSAFRVEV 114
              |  ||                    | :|  |     |||        |:
Db        261 DLEDS-----------------------FTILQSKDLKQEPLD------DPTCI---- 285

Qy        115 QADGTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDI 174
                | :  ::  :| |   : |: |||  |  || |  |                 || |
Db        286 --DTSETSLSNQNKLFSDINLNDQEWQEL--ID-ELAN---------------TVPEDD 324

Qy        175 AEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKF 234
              :         | |    : |: :                    :||  |  ::||:|||
Db        325 IQ---------DLFNEDFEEKKEPEFSQ-----------PATETPLSQESASVKSDPSH 363

Qy        235 ----WVSVGAIAA--GLAGLAATGIAQALAL-----TPEPDDPTTTDPDQAANAAESATK 283
                  ||:|:  |   :|   : :: | :|     ||   :| ::  :  |  : ::  :
Db        364 SPFAHVSMGSPQARPSSSGPPFSTVSTATSLPSVASTPAAPNPASSPANCAVQSPQTPNQ 423

Qy        284 DQLTQEA--------FKNPENQKVNIDANGN-AIPSGEL--XDDIVEQIAQQAKEAGEVA 332
                :|          ||     |   |:|  |:|| ::   : : :  ||| : |  :
Db        424 AHTPGQAPPRPGNGYLLNPAAVTVAGSASGPVAVPSSDMSPAEQLKQMAAQQQQRAKLMQ 483

Qy        333 RQQAVESNAQAQQRYEDQHARRQEELQLSS 362
              ::|  :   | ||: : | .::|:: | |:
Db        484 QKQQQQQQQQQQQQQQQQQQQQQQQQQHSN 513


RESULT 12
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12713
;   LENGTH: 2344
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12713

  Query Match             4.8%;  Score 135;  DB 9;  Length 2344;
  Best Local Similarity   19.2%;  Pred. No. 0.31;
  Matches  107;  Conservative  102;  Mismatches  221;  Indels  126;  Gaps   23;

Qy         24 SQTDGAARGGTGHLISSTGAL---GSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAAT 80
              : |  :    | : ||:: :|   || |    : ||:::|| |      | :  : :: :|
Db        905 ASTSQSVSDSTSNSISTSESLSESGSTSESISISNSISNSV-SASTSKLESQSTSISLST 963

Qy         81 SETCLLGGFEVLHD--------KGPLDI-----LNTQIGPSAFRVEVQAD-----GTHAA 122
              |::  :  | | |         | | :     ::|    |    |: :|     |: ||
Db        964 SDSKSMSTSESLSDSTSTSDSVSGSLSVAGSQSVSTSTSDSMSTSEMISDSMSTSGSLAA 1023

Qy        123 IGEKNGLEVSVTLSPQE--------WSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDI 174
                 |: : || ::|  :          |: : |:: |:  : |  : ||     : :| :
Db       1024 SDSKS-MSVSSSMSTSQSGSTSESLSDSISTSDSDSKSLSLST-SQSGSTSTSTSTSSSV 1081

Qy        175 AEARTRILAKLDPDNHGGRQPKDVDTRSVGVG---------SASGIDDGVVSETHTSTTN 225
                : ::           |       |: |:            |||      :|:: :::|:
Db       1082 RMSESQ-------STSGSMSTSQSDSTSISTSFSDSTSDSKSASTASSESISQSVSTSTS 1134

Qy        226 SSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQ 285
               || :          :::   :   :| :: : :|:    | |:             ::| |
Db       1135 GSVSTST-------SLSTSNSERTSTSMSDSTSLSTSESDSTSD---------STSTSDS 1178

Qy        286 LTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQ 345
              :: ||   |:  :::  : :   |              ::| |      :   || :::
Db       1179 IS-EAISGSESTSISLSESNSTSDS-------------ESKSASAFLSESLSESTSES-- 1222

Qy        346 RYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTT----- 400
                       | | |:     || :   :|      :|:|      :   :|:|
Db       1223 --------TSESLSGSTSDSTSLSDSNSESG----STSTSLSNSTSGSASISTSTSGSAS 1270

Qy        401 THTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGAR 460
              | ||  :          :|      ||      :  ||  |||:|          |
Db       1271 TSTVKSESV------STSLSTSTSTSLSD-----STSLSTSLSDSTS---------GSKS 1310

Qy        461 NSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLR 520
              |||||         | ::     |:  : | ||  :          |  |:    ::|  :
Db       1311 NSLSASMSTS---DSISTRKSESLSASTSLSGSTS-----ESESGSTSSSESKSDSTSMS 1362

Qy        521 LGMGGLTSGGETAVSS 536
              | |   || | |:||:
Db       1363 LSMSQSTS-GSTSVST 1377


RESULT 13
US-10-156-761-12878
; Sequence 12878, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
```

```
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12878
;   LENGTH: 864
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-12878

  Query Match             4.8%;  Score 133;  DB 15;  Length 864;
  Best Local Similarity   21.5%;  Pred. No. 0.11;
  Matches  124;  Conservative   84;  Mismatches  191;  Indels  178;  Gaps   32;

Qy          5 NLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALG-SRSLFSPLRNSMADSVD- 62
              |  |     |       ||  |    |   ||:| |  |   :||   ||  | : :  ||
Db        353 NFANPAGAN------PPALSALPGL--GGSGGL-SQLSSLGIGASLDFPGQFARFDSAPV 403

Qy         63 SRD--IPGLPTNPSRLAAATSETCLLGGFEVLHDKGP---LDILNTQIGPSAFRVEVQAD 117
              :||  | | ||    : : : :| | |    ::| ||     :| :|:  :  |||
Db        404 TRDLRISGSPTATVHVKSTSEDTVLFG---KVYDVGPGGTQKVLPSQL-VTPIRVE---- 455

Qy        118 GTHAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEA 177
                  | | | :|::        || :|| | :           |||
Db        456 ------GAKEGKDVTL--------SLPAIDHEVR----------SGH------------ 478

Qy        178 RTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSD------ 231
              | |::                          :  | |  | |  :|   |:: |
Db        479 RLRLV------------------------LASTDLGYASPTAPATYTVSLKGDLLVPTA 513

Qy        232 PKFWVSVGAIAAGLAGLAATGIAQALAL---------TPEPDDPTTTDP----DQAANAA 278
              |    :   : | :  |   |:  ||||         || ||      |    | :   |
Db        514 PGVTTAAAPLPAWVWWLPLAGVVVALALLLSGRRRTATPAPDPALAEVPLEITDLSKRYA 573

Qy        279 ESATKDQLTQEAFKNPENQKVNI--------------------DAN-----GNAIPSGE 312
              . :|| :  :   :|:  : | : :                ||      |:|| |
Db        574 KSADRYAVRDLSFRVEKGQVLGLLGPNGAGKTTTLRMLMGLIKPDAGEIHVFGHAIRPGA 633

Qy        313 LXDDIVEQIAQQAKEAG---EVARQQAVESNAQAQQR-YEDQHARRQEELQLSSGIGYGL 368
                 :: ::    : ||    :: :: :|   ||  |  || |    | ||: :|:|  |
Db        634 ---PVLSRVGSFVEGAGFLPHLSGRENLELYWQATGRPSEDAHL--DEALQI-AGLGDAL 687

Qy        369 SSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFS 428
              : |:          :  : :|   |:         :: : | |:   ::    :| |
Db        688 ARAV-------RTYSQGMRQRLAIAQAMLGLPDLLILDEPTNGLDPPQI----REMREVM 736

Qy        429 DRRDSQGS--VASTHW----SDSSSEVV----------NPYAEVGGARNSL---SAHQPE 469
              | : |   : |:|       | : :|           |  |: |: ::|   :    :
Db        737 IRYAAAGRTVIVSSHLLAEVEQSCTHLVVMDRGRLVQAGPVHEIVGSGDTLLVGTGTPVD 796

Qy        470 EHIYDEVAADPGY-SVIQNFSG------SGPVTGRLI 499
              | : ::| | ||  | ::   |       :    |||:
Db        797 EPVVEKVGALPGVASAVRTDEGLLVRLDTDGSAGRLV 833


RESULT 14
US-09-379-931-7
; Sequence 7, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
;  APPLICANT:  Smit, John
;  APPLICANT:  Bingle, Wade H.
;  APPLICANT:  No. US20020009792A1ellini, John F.
;  TITLE OF INVENTION:  EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULC
;  FILE REFERENCE:  08106/002003
;  CURRENT APPLICATION NUMBER: US/09/379,931
;  CURRENT FILING DATE:  1999-08-24
;  PRIOR APPLICATION NUMBER:  US 08/614,377
;  PRIOR FILING DATE:  1996-03-12
;  PRIOR APPLICATION NUMBER:  US 08/194,290
;  PRIOR FILING DATE:  1994-02-09
;  PRIOR APPLICATION NUMBER:  US 07/895,367
;  PRIOR FILING DATE:  1992-06-09
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1026
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-379-931-7

  Query Match             4.7%;  Score 132.5;  DB 9;  Length 1026;
  Best Local Similarity   21.8%;  Pred. No. 0.15;
  Matches  123;  Conservative   64;  Mismatches  222;  Indels  155;  Gaps   27;

Qy         26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCL 85
              || ||         |:|  ||         |:    |:  : |   | : :|   :
Db        230 TDNAAGVNLFTAYPSSGVSGSTL-------SLTTGTDT--LTGTANNDTFVAGEVAGAAT 280

Qy         86 LGGFEVLHDKGPLDILN----------TQIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134
              |   : |       |:||          | :  |          |    :   :|:
Db        281 LTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSGVTGLTA 340

Qy        135 LSPQEWSSLQSIDT-EGKNRFVFTGGRGGS-----GHPMVTVASDIAEARTRILAKLDPD 188
              |:     : |::    |:|    |  :      |   |||||       | : :
Db        341 LNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS------TGVTS----- 389

Qy        189 NHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGL 248
                        | :||  ||:    | ||    |  |||         : ||||  : |
Db        390 ----------GTTTVGANSAA---SGTVS---VSVANSST-------TTTGAIA--VTGG 424

Qy        249 AATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAI 308
              |  :||          : | |  |        | |   :|| |             |  |
Db        425 TAVTVAQTAG---NAVNTTLTQADVTVTGNSSTTAVTVTQTA----------AATAGAT 470

Qy        309 PSGELXD--DIVEQIAQQAKEAGEVAR-------QQAVESNAQAQQRYEDQHARRQEELQ 359
              :| :     | :  |  |  ||::|          ::|:|
Db        471 VAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALTTVNL---------- 518

Qy        360 LSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVAL 419
                 | |  |         |||  |  ||      |   | |   : :    || |   : |
Db        519 --SGTGTSL--------GIGRGALTA-----TPTANTLTLNVNGLT--TTGAITDSEAA- 560

Qy        420 MPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEV----GGARNSLSAHQPEEHIYDE 475
                       :|   :  ::| :  | : : :|    |       | || ::::|
Db        561 --------ADDGFTTINIAGSTASSTIASLVAADATTLNISGDARVTITSH--------T 604

Qy        476 VAADPGYSVIQNFS---GSGPVTGRLIGTPGQGIQS------TYALLANSG-------GL 519
              ||  | :|  :      |:   || |: | | |  |       | |::  :|
Db        605 AAALTGITVTNSVGATLGAELATG-LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSA 663

Qy        520 RLGMGGLTSGG---ETAVSSVNAA 540
              || ||  :||   :  |::|| :
Db        664 TLGAGGSVNGGDGTDVLVANVNGS 687


RESULT 15
US-10-223-597-7
; Sequence 7, Application US/10223597
; Publication No. US20030135037A1
; GENERAL INFORMATION:
;  APPLICANT: Smit, John
;  APPLICANT:  Bingle, Wade H.
;  APPLICANT:  No. US20030135037A1ellini, John F.
;  TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOB
;  FILE REFERENCE: 08106/002003
;  CURRENT APPLICATION NUMBER: US/10/223,597
;  CURRENT FILING DATE:  2002-08-19
;  PRIOR APPLICATION NUMBER: US/09/379,931
;  PRIOR FILING DATE: 1999-08-24
;  PRIOR APPLICATION NUMBER: US 08/614,377
;  PRIOR FILING DATE: 1996-03-12
;  PRIOR APPLICATION NUMBER: US 08/194,290
;  PRIOR FILING DATE: 1994-02-09
;  PRIOR APPLICATION NUMBER: US 07/895,367
```

```
;  PRIOR FILING DATE: 1992-06-09
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;    LENGTH: 1026
;    TYPE: PRT
;    ORGANISM: Caulobacter crescentus
US-10-223-597-7

  Query Match             4.7%;  Score 132.5;  DB 12;  Length 1026;
  Best Local Similarity   21.8%;  Pred. No. 0.15;
  Matches  123;  Conservative   64;  Mismatches  222;  Indels  155;  Gaps   27;

Qy         26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCL 85
              || ||          |:|  ||          |:    |:  : |   | : :|   :
Db        230 TDNAAGVNLFTAYPSSGVSGSTL-------SLTTGTDT--LTGTANNDTFVAGEVAGAAT 280

Qy         86 LGGFEVLHDKGPLDILN-----------TQIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134
              |   : |       |:||           | :  |          |     :   :|:
Db        281 LTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSGVTGLTA 340

Qy        135 LSPQEWSSLQSIDT-EGKNRFVFTGGRGGS-----GHPMVTVASDIAEARTRILAKLDPD 188
              |:     : |::    |:|    |  :  :     |   |||||      | : :
Db        341 LNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS------TGVTS----- 389

Qy        189 NHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGL 248
                        | :||  ||:    | ||    |  |||          : ||||  : |
Db        390 ----------GTTTVGANSAA---SGTVS---VSVANSST-------TTTGAIA--VTGG 424

Qy        249 AATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAI 308
              |  :||          : | |. |       | |   :|| |           |   |
Db        425 TAVTVAQTAG---NAVNTTLTQADVTVTGNSSTTAVTVTQTA-----------AATAGAT 470

Qy        309 PSGELXD--DIVEQIAQQAKEAGEVAR-------QQAVESNAQAQQRYEDQHARRQEELQ 359
              :| :      | :  |  |  ||::|          ::|:|
Db        471 VAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALTTVNL------------ 518

Qy        360 LSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVAL 419
                || |  |        ||| |  ||       |   | |   : :    || |   : |
Db        519 --SGTGTSL--------GIGRGALTA-----TPTANTLTLNVNGLT--TTGAITDSEAA- 560

Qy        420 MPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEV----GGARNSLSAHQPEEHIYDE 475
                    :|   :  ::| :  | : : :|   |       | || ::::|
Db        561 --------ADDGFTTINIAGSTASSTIASLVAADATTLNISGDARVTITSH--------T 604

Qy        476 VAADPGYSVIQNFS---GSGPVTGRLIGTPGQGIQS------TYALLANSG-------GL 519
              ||  | :|  :     |:   || |: | | |  |       | |::  :|
Db        605 AAALTGITVTNSVGATLGAELATG-LVFTGGAGRDSILLGATTKAIVMGAGDDTVTVSSA 663

Qy        520 RLGMGGLTSGG---ETAVSSVNAA 540
              || ||  :||   :  |::|| :
Db        664 TLGAGGSVNGGDGTDVLVANVNGS 687


Search completed: December 16, 2003, 09:07:33
Job time : 28.7563 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 25.7653 Seconds
(without alignments)
901.547 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	149	5.3	831	4	US-09-336-447A-1	Sequence 1, Appli
2	144	5.1	3892	4	US-09-328-352-5503	Sequence 5503, Ap
3	141	5.0	580	2	US-08-672-564-1	Sequence 1, Appli
4	139.5	5.0	2354	4	US-09-268-347-47	Sequence 47, Appl
5	139.5	5.0	2411	4	US-09-268-347-36	Sequence 36, Appl
6	138.5	4.9	2353	3	US-09-377-155-33	Sequence 33, Appl
7	138.5	4.9	2353	3	US-08-913-942-4	Sequence 4, Appli
8	138.5	4.9	2353	4	US-09-669-974-33	Sequence 33, Appl
9	137	4.9	2736	4	US-09-252-991A-30227	Sequence 30227, A
10	134.5	4.8	734	4	US-09-328-352-4412	Sequence 4412, Ap
11	134.5	4.8	1912	1	US-08-409-995-4	Sequence 4, Appli
12	134.5	4.8	1912	3	US-08-685-467-4	Sequence 4, Appli
13	134	4.8	375	4	US-09-252-991A-21021	Sequence 21021, A
14	133	4.8	2703	1	US-08-185-432-19	Sequence 19, Appl
15	133	4.8	2703	4	US-08-899-232-4	Sequence 4, Appli
16	132.5	4.7	1026	1	US-08-194-290-7	Sequence 7, Appli
17	132.5	4.7	1026	2	US-08-614-377A-7	Sequence 7, Appli
18	132.5	4.7	1026	3	US-09-142-648B-7	Sequence 7, Appli
19	132.5	4.7	5215	3	US-09-105-537-2	Sequence 2, Appli
20	132	4.7	1415	4	US-09-252-991A-26438	Sequence 26438, A
21	128	4.6	1507	6	5268270-2	Patent No. 5268270
22	126	4.5	376	4	US-09-056-556-202	Sequence 202, App
23	126	4.5	376	4	US-09-072-596-197	Sequence 197, App
24	126	4.5	444	4	US-09-252-991A-21658	Sequence 21658, A
25	126	4.5	646	4	US-09-252-991A-23299	Sequence 23299, A
26	125	4.5	851	4	US-09-252-991A-22021	Sequence 22021, A
27	125	4.5	2442	4	US-09-514-247A-10	Sequence 10, Appl
28	124.5	4.4	729	4	US-09-223-040-2	Sequence 2, Appli
29	124.5	4.4	1034	4	US-09-252-991A-26658	Sequence 26658, A
30	124	4.4	925	4	US-09-252-991A-27057	Sequence 27057, A
31	124	4.4	1180	4	US-09-206-942-65	Sequence 65, Appl
32	124	4.4	1188	4	US-09-206-942-63	Sequence 63, Appl
33	124	4.4	1536	4	US-09-206-942-67	Sequence 67, Appl
34	123.5	4.4	665	4	US-09-252-991A-20383	Sequence 20383, A
35	122.5	4.4	507	3	US-08-860-635A-19	Sequence 19, Appl
36	122.5	4.4	507	4	US-09-281-476-19	Sequence 19, Appl
37	122	4.4	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
38	121	4.3	608	4	US-08-843-572E-2	Sequence 2, Appli
39	121	4.3	1638	4	US-09-071-035-258	Sequence 258, App
40	121	4.3	1638	4	US-09-071-035-262	Sequence 262, App
41	121	4.3	1638	4	US-09-071-035-266	Sequence 266, App
42	120.5	4.3	2039	4	US-09-077-098A-7	Sequence 7, Appli
43	120	4.3	2042	4	US-09-077-098A-6	Sequence 6, Appli
44	119.5	4.3	729	4	US-09-625-188-20	Sequence 20, Appl
45	118.5	4.2	1287	4	US-09-252-991A-29606	Sequence 29606, A

ALIGNMENTS

RESULT 1
US-09-336-447A-1
; Sequence 1, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-1

Query Match 5.3%; Score 149; DB 4; Length 831;
Best Local Similarity 19.2%; Pred. No. 0.00027;
Matches 101; Conservative 73; Mismatches 193; Indels 160; Gaps 21;

```
Qy        74 SRLAAATSETCLLGGFEVLHDKGPLDILNTQIGP---SAFRVEVQADGTHAAIGEKNGLE 130
             |     || : |||       |    : | ::|    : |  | :|:    ::| : |
Db       226 SGTTGVTSNSVLLGNETA--GKQATTVKNAEVGGLSLTGFAGESKAENGVVSVGSEGGER 283

Qy       131 VSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVAS----DIAEARTRILAKLD 186
               | :       :    |: |         |   | : ||      ||   |  ||   |
Db       284 QIVNVGA---GQISDTSTDAVN--------GSQLHALATVVDDNQYDIVNNRADILNNQD 332

Qy       187 P-----------DNHGGRQPKDVDT----------------------RSVGVGS----- 207
                       ||  |   :|:::                          ||| |
Db       333 DIKDLQKEVKGLDNEVGELSRDINSLHDVTDNQQDDIKELKRGVKELDNEVGVLSRDINS 392

Qy       208 -----ASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPE 262
                  |   ||   ::       |  |:   |    || ::  :  |
Db       393 LHDDVADNQDDIAKNKADIKGLNKEVKELDK---EVGVLSRDIGSL-------------- 435

Qy       263 PDDPTTTDPDQAANAAESATKDQLTQEAFKNPE----NQKVNIDAN-------------- 304
               ||  |   | | | |:  | :   :|    |       :|| :|| |
Db       436 HDDVATNQADIAKNQADIKTLENNVEEELLNLSGRLLDQKADIDNNINNIYELAQQQDQH 495

Qy       305 -------GNAIPSG--ELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQ 355
                    | :  |  :|   :::| |  ||   ::|: |    :  |    :||:|::|
Db       496 SSDIKTLKNNVEEGLLDLSGRLIDQKADIAKNQADIAQNQTDIQDLAAYNELQDQYAQKQ 555
```

```
Qy        356 EE--------------------LQLSSGIG--YGL-------SSALIVAGGIGAGVTTAL 386
Db        556 TEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKVSAANTDRI 615

Qy        387 HRRNQPAE---QTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWS 443
Db        616 AKNKAEADASFETLTKNQNTLIEQGEALVEQNKA--INQELEGFAAHADIQ--------- 664

Qy        444 DSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSG 490
Db        665 --DKQILQNQADITTNKTAI-----EQNINRTVA--NGFEIEKNKAG 702


RESULT 2
US-09-328-352-5503
; Sequence 5503, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5503
;   LENGTH: 3892
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5503

  Query Match             5.1%;  Score 144;  DB 4;  Length 3892;
  Best Local Similarity   19.6%;  Pred. No. 0.0082;
  Matches  113;  Conservative   80;  Mismatches  253;  Indels  130;  Gaps   21;

Qy          3 IGNLGNNVNGNHLIPPAPPLPSQTDGAARG----GTGHLISSTGALGSRSLFSPLRNSMA 58
Db        178 IAAASNHDSNDDSSAPADTTPPSTDGVTFSVDPVTSDNVINASEASGNVTITGVLKNIPA 237

Qy         59 DSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
Db        238 D-----------------AANTAVTVVINGV----------TYNATVDKAAGTWTVSVPG 270

Qy        119 THAAIGEKNGLEVSVTLSPQEWSS-------LQSIDTEGKNRFVFTGGRG-----GSGHP 166
Db        271 SGLVADADKTIDAKVTFTDAAGNSSSVNDTQTYTLDTTAPNAPVIDPVNGTDPITGTAEP 330

Qy        167 MVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNS 226
Db        331 GSTVTVTYPDGSTKTVV-AGPDG----------TWTV---PNPGLNDG--------DEVT 368

Qy        227 SVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQL 286
Db        369 AVATDP----------AGNTSGPATAVVDAVAPTVALDDVLTND----STPALTGTVNDP 414

Qy        287 TQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQ--------QAKEAGEVARQQAVE 338
Db        415 TATVVVNVDG--VDYPAVNNGDGTWTLADNTLPTLADGPHTITVTATDAAGNVGTDTGVV 472

Qy        339 SNAQAQQRYEDQHARRQEELQLSSGIGYGLSSA----LIVAGGIGAGVTTALHRRNQPAE 394
Db        473 TVDTAAPN--------------TAGVTFTIDSVTADNVINASEAAGNVTITGVLKNIPAD 518

Qy        395 QTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYA 454
Db        519 ATNTAVTVVI-----NGVTYN--ATVDKTAGTWTVSVPGSGLVAD---ADKTIDAKVTFT 568

Qy        455 EVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTY---- 510
Db        569 DAAGNSSTVN----DTQIYTLDTAAPAAPVIDPVNGTDPITG--TAEPGSTVTVTYPNGD 622

Qy        511 ---ALLANSGGLRLGMGGLTSGGETAVSSVNAAPTP 543


                  :    |  :   ||  | |    : : | |
Db        623 TATVVAGPDGSWSVPNPGLNDGDEVEAIATDPAGNP 658


RESULT 3
US-08-672-564-1
; Sequence 1, Application US/08672564
; Patent No. 5824503
;  GENERAL INFORMATION:
;    APPLICANT:  KUROME, Yoko
;    APPLICANT:  IZU, Hiroyuki
;    APPLICANT:  IZUMI, Yoshiya
;    APPLICANT:  SANO, Mutsumi
;    APPLICANT:  KATO, Ikunoshin
;    APPLICANT:  ITO, Makoto
;    TITLE OF INVENTION:  GENE ENCODING ENDOGLYCOCERAMIDASE ACTIVATOR
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch, LLP
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/672,564
;      FILING DATE:  28 JUNE 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WEINER, Marc S.
;      REGISTRATION NUMBER:  32,181
;      REFERENCE/DOCKET NUMBER:  1422-0263P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  580
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-672-564-1

  Query Match             5.0%;  Score 141;  DB 2;  Length 580;
  Best Local Similarity   23.9%;  Pred. No. 0.00076;
  Matches   84;  Conservative   45;  Mismatches  126;  Indels   96;  Gaps   16;

Qy          4 GNLGNNVNGNHLIPP----------APPLPSQTDGAARGGTGHLISSTGALGSRSLFSPL 53
Db        215 GQFGNDANFLTFLPRASAPIVGTVWAPTQCSPRDTAA----GPLNAGAGPLATIQI---L 267

Qy         54 RNSMADSVDSRDIPGLPTNPSR--LAAATSETCLLGGFEVLHD----KGPLDILNTQIGP 107
Db        268 RQAVA-TVSYLDGPSAVTNGGEFTLNATVVPTPDSGQVQFTRDGEDVGAPVDLVN---GK 323

Qy        108 SAFRVEVQADGTHAAIGEKNGLEV--------SVTLSPQEWSSLQSIDTEGKNRFVFTGG 159
Db        324 ASLTQSLDTDGDYAYEAKFLGAEFFNPSSAAKTVTVTSQDIQTTTSVT------------ 371

Qy        160 RGGSGHPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSET 219
Db        372 --GPDH------DAYRDQPVNLTAKVEPGVSGGTVAFEVD--GTPVGTADVMDDGAAVLP 421

Qy        220 HTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALAL--------TPEPDDPTTTDP 271
Db        422 HTFTTNGTHR-----------VIARYSG--AEGISPSVSLQYPVSVTEAPAADVATTITV 468
```

```
Qy        272 DQAANAAESATKDQLTQEAFKNPENQK---------------VNIDANGNA 307
              |  |: |: :    :|  |  :| : :                | :|||| |
Db        469 DPIASTAKGS---PVTLTARLDPADARGTVQFKLGDVLLGGPVRVDANGVA 516


RESULT 4
US-09-268-347-47
; Sequence 47, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 2354
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-47

  Query Match             5.0%;  Score 139.5;  DB 4;  Length 2354;
  Best Local Similarity   22.0%;  Pred. No. 0.0093;
  Matches  124;  Conservative   63;  Mismatches  206;  Indels  171;  Gaps   26;

Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
              ||   ||                    |  || :|:  |      |   | |  |   |: :
Db       1163 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1198

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
                    : | :    |  |: : |  |    | ||          | |    ||
Db       1199 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1243

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
              |:  ||  ||                  |||   :    :|   ||:  |:|||:
Db       1244 AV--KNANEVEFV--------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1284

Qy        179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
                  |   ||  ||  |      ||  : |   | |  : | ::  |:  ||:: |
Db       1285 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1343

Qy        230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
                        :| |   :|| :|  :   ||| :        | ||  | |     :
Db       1344 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1396

Qy        278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
              |  | |  | || :| |   | ||  |          ||  :: |  :  | :
Db       1397 ANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1444

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
               |     :  : |: :  :  :|         : | :||    |:  |      : :| || |
Db       1445 LGTNGNKVNITSDTKGLKFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
                                  ||         :   |::| :  :|
Db       1498 ----------------------GG-----NGITDNEKKRAASVKD--------------- 1515

Qy        448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
               |:|    | | : : ||:   |:| | ||    |  :   ||    |   : :   | :
Db       1516 -VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1569

Qy        508 STYALLANSGGLRLGMGGLTSGGE 531
              :   : | :  ::    | | :| |
Db       1570 TEVKIGAKTSVIKDHNGKLFTGKE 1593


RESULT 5
US-09-268-347-36
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 2411
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-36

  Query Match             5.0%;  Score 139.5;  DB 4;  Length 2411;
  Best Local Similarity   21.5%;  Pred. No. 0.0096;
  Matches  125;  Conservative   83;  Mismatches  228;  Indels  145;  Gaps   28;

Qy          6 LGNNVNGNH----LIPPAPPLPSQTDGAARG---GTGHLISST--GALGSRSLFSPLRNS 56
              ||   || :    :|       :  :||| |      |: || |  |         : ::::
Db       1161 LGGTANGRNDTGTVINKDGLTITLANGAAAGTDASNGNTISVTKDGISAGNKEITNVKSA 1220

Qy         57 MADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA 116
              :    |::: |  | |   || |:|            :  :|:     | :    : :|
Db       1221 LKTYKDTQNTAG-ATQP---AANTAEVA---------KQDLVDLTKPATGAAGNGADAKA 1267

Qy        117 -DGTHAAIGEKNGL-------EVSVTLSPQEW----SSLQSIDTEGKNRFVFTGGRGGSG 164
               | | | :|:  ||       : :     :|:     :   ::  |||    :     :|
Db       1268 PDTTAATVGDLRGLGWVLSAKKTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNG 1327

Qy        165 HPMVTVASDIAEAR--------TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVV 216
               ||:  |:|||:         |    ||  ||  |      ||  : |   | |  : |
Db       1328 KHTVTI--DVAEAKVGDGLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVT 1384

Qy        217 SETHTST-TNSSVR--------------SDPKFWVSVGAIAAGLAGLAATGIAQALALT 260
              ::  |:  ||:: |              :| |   :|| :|  :   ||| :
Db       1385 TDATTAQGTNANERGKVVVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------K 1437

Qy        261 PEPDDPTTTDPDQAANAAESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPS 310
              | ||  | |     : |  | |  | || :| |   | ||  |          ||  :: |
Db       1438 VENDDSATIDDSPTDDGANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKS 1494

Qy        311 GELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSS 370
               : | :          |   :  : |: :    :|         : | :||    |:
Db       1495 ATVSDKL---------SLGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTD 1544

Qy        371 ALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDR 430
              |      : :| || |                  ||       :   |::| :
Db       1545 TL-----LNSGATTNL----------------------GG-----NGITDNEKKRAASV 1571

Qy        431 RDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSG 490
               :|                |:|    | | : : ||:   |:| | ||    |  :   ||
Db       1572 KD----------------VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSG 1610

Qy        491 SGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGE 531
                  |   : :   | ::   : | :  ::    | | :| |
Db       1611 DKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHNGKLFTGKE 1651


RESULT 6
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
;  APPLICANT:  PEAK, Ian Richard Anselm
;  APPLICANT:  JENNINGS, Michael Paul
;  APPLICANT:  MOXON, E. Richard
;  TITLE OF INVENTION:  NOVEL SURFACE ANTIGEN
;  FILE REFERENCE:  065064/0128
;  CURRENT APPLICATION NUMBER: US/09/377,155
;  CURRENT FILING DATE:  1999-08-19
;  PRIOR APPLICATION NUMBER:  PCT/AU98/01031
;  PRIOR FILING DATE:  1998-12-14
;  PRIOR APPLICATION NUMBER:  GB 9726398.2
```

```
;  PRIOR FILING DATE:  1997-12-12
;  NUMBER OF SEQ ID NOS:  33
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-377-155-33

  Query Match             4.9%;  Score 138.5;  DB 3;  Length 2353;
  Best Local Similarity   22.0%;  Pred. No. 0.011;
  Matches  124;  Conservative   62;  Mismatches  207;  Indels  171;  Gaps   26;

Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
              ||   ||                  |  || :|:  |      |   | | |   |: :
Db       1163 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1198

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
                    : | :    |  |: : |  |    | ||         | |    ||
Db       1199 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1243

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
              |:  ||  ||                  |||    :     :|   ||:  |:|||:
Db       1244 AV--KNANEVEFV--------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1284

Qy        179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
                  |    ||  ||  |    ||  : |   | |  : | ::  |:  ||:: |
Db       1285 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1343

Qy        230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
                       :| |   :|| :|  :   ||| :        | ||  | |     :
Db       1344 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1396

Qy        278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
              |  | |  | || :| |   | ||  |         ||  :: |  : | :
Db       1397 ANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1444

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
                |   :  :|: :    :|        : |:||   |:  |      : :| || |
Db       1445 LGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
                                  ||       :   |::| :  :|
Db       1498 ----------------------GG-----NGITDNEKKRAASVKD--------------- 1515

Qy        448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
               |:|    | | : :||:   |:| | ||    |  :  ||    |  : :   | :
Db       1516 -VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1569

Qy        508 STYALLANSGGLRLGMGGLTSGGE 531
              :   : | :  ::   | | :| |
Db       1570 TEVKIGAKTSVIKDHNGKLFTGKE 1593


RESULT 7
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme, Joseph
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/913,942
;      FILING DATE:  29-DEC-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/4031
;      FILING DATE:  22-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vance, Dolly A.
;      REGISTRATION NUMBER:  39,054
;      REFERENCE/DOCKET NUMBER:  A-61053-1/RFT/RMS/DAV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2353 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-913-942-4

  Query Match             4.9%;  Score 138.5;  DB 3;  Length 2353;
  Best Local Similarity   22.0%;  Pred. No. 0.011;
  Matches  124;  Conservative   62;  Mismatches  207;  Indels  171;  Gaps   26;

Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
              ||   ||                  |  || :|:  |      |   | | |   |: :
Db       1163 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1198

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
                    : | :    |  |: : |  |    | ||         | |    ||
Db       1199 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1243

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
              |:  ||  ||                  |||    :     :|   ||:  |:|||:
Db       1244 AV--KNANEVEFV--------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1284

Qy        179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
                  |    ||  ||  |    ||  : |   | |  : | ::  |:  ||:: |
Db       1285 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1343

Qy        230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
                       :| |   :|| :|  :   ||| :        | ||  | |     :
Db       1344 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1396

Qy        278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
              |  | |  | || :| |   | ||  |         ||  :: |  : | :
Db       1397 ANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1444

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
                |   :  :|: :    :|        : |:||   |:  |      : :| || |
Db       1445 LGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
                                  ||       :   |::| :  :|
Db       1498 ----------------------GG-----NGITDNEKKRAASVKD--------------- 1515

Qy        448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
               |:|    | | : :||:   |:| | ||    |  :  ||    |  : :   | :
Db       1516 -VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1569

Qy        508 STYALLANSGGLRLGMGGLTSGGE 531
              :   : | :  ::   | | :| |
Db       1570 TEVKIGAKTSVIKDHNGKLFTGKE 1593
```

```
RESULT 8
US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
;  APPLICANT: PEAK, Ian Richard Anselm
;  APPLICANT:  JENNINGS, Michael Paul
;  APPLICANT:  MOXON, E. Richard
;  TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
;  FILE REFERENCE: 065064/0128
;  CURRENT APPLICATION NUMBER: US/09/669,974
;  CURRENT FILING DATE:  2000-09-26
;  PRIOR APPLICATION NUMBER: US 09/377,155
;  PRIOR FILING DATE: 1999-08-19
;  PRIOR APPLICATION NUMBER: PCT/AU98/01031
;  PRIOR FILING DATE: 1998-12-14
;  PRIOR APPLICATION NUMBER: GB 9726398.2
;  PRIOR FILING DATE: 1997-12-12
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-669-974-33

  Query Match             4.9%;  Score 138.5;  DB 4;  Length 2353;
  Best Local Similarity   22.0%;  Pred. No. 0.011;
  Matches  124;  Conservative   62;  Mismatches  207;  Indels  171;  Gaps   26;

Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
              ||   ||                    |  || :|:  |       |   | |  |   |: :
Db       1163 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1198

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
                        : | :    |  |: : |  |     |  ||          | |    ||
Db       1199 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1243

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
              |:  ||  ||                  |||    :     :|   ||:  |:|||:
Db       1244 AV--KNANEVEFV--------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1284

Qy        179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
                  |    ||  ||  |      ||  :|   | |  : | ::  |:  ||:: |
Db       1285 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1343

Qy        230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
                        :| |   :|| :|  :   ||| :         | ||  | |     :
Db       1344 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1396

Qy        278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
              |  | |  | || :| |   | ||  |          ||  :: |  : | :
Db       1397 ANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1444

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
               |    :  : |: :     :|         : | :||    |:  |     : :| || |
Db       1445 LGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
                                  ||        :   |::| :  :|
Db       1498 ----------------------GG-----NGITDNEKKRAASVKD--------------- 1515

Qy        448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
               |:|    | | : : ||:   |:| | ||    |  :   ||    |   : :   | :
Db       1516 -VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1569

Qy        508 STYALLANSGGLRLGMGGLTSGGE 531
              :   : | :  ::    | | :| |
Db       1570 TEVKIGAKTSVIKDHNGKLFTGKE 1593


RESULT 9
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 30227
;   LENGTH: 2736
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

  Query Match             4.9%;  Score 137;  DB 4;  Length 2736;
  Best Local Similarity   21.5%;  Pred. No. 0.019;
  Matches  135;  Conservative   56;  Mismatches  275;  Indels  162;  Gaps   28;

Qy          2 PIGNLGNNVNGNHLIPPAPPLPSQT-------DGAARGGTGHLISSTGALGSRSLFSPLR 54
              |||    : |||    |   ||  |       | |         |:  |   :  :  |
Db        477 PIGQTSADANGNWSFTPGSQLPDGTVVNVVARDAAGNSSPATSITVDGVAPNAPVVEPSN 536

Qy         55 NS-------MADSVDSRDIPGLPTNPSRLAA-----ATSETCLLGGFEVLHDKGPLDILN 102
                       |          ||   | | |   :   |       | | | |  |  :    |
Db        537 GSELSGTAEPGSSVTLTDGNGNPIGQTTADANGNWSFTPSTPLPDGTVV--NVVARDAAG 594

Qy        103 TQIGPSAFRVEVQADGTHAAIGEKNGLEVS--------VTLSPQEWSSLQSIDTEGKNRF 154
                  |::  |:  | |   :   ||  :|        |||:   : :  :  :|   :
Db        595 NSSPPASVTVDAVAPAT-PTVDPSNGTTLSGTAEPGATVTLTDGNGNPIGQVTADGSGNW 653

Qy        155 VFTGG----------------RGGSGHPMVTVASDIAEARTRILAKLDPDN----HGGRQ 194
              ||                   | :  |       :| |   :    :| |      |  :
Db        654 TFTPSTPLPNGTVVNATATDPSGNASSPASVTVDPVAPATPVV----NPSNGTTLSGTAE 709

Qy        195 PKDVDTRSVGVGSASG--IDDGVVSETHTSTT---NSSVRSDPKFWVSVGAIAAGLAGLA 249
              |    | : | |:  |    ||  : : | ||   | :|         | | |   :|
Db        710 PGATVTLADGNGNPIGQVTADGSGNWSFTPTTPLPNGTV---------VNATATDASGNT 760

Qy        250 ATG----IAQALALTP--EPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDA 303
              : |    :      ||   | : ||      :  ||  :   ||  :   ||  | |  |
Db        761 SAGSSVTVDSVAPATPVINPSNGTT-----LSGTAEPGSSVTLT-DGNGNPIGQ-VTADG 813

Qy        304 NGN--AIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLS 361
              :||    ||  | | |   |      ||  : | :             | |      : ||
Db        814 SGNWSFTPSTPLADGTVVN-ATATDPAGNTSGQGSTT---------VDGVAPTTPTVNLS 863

Qy        362 SGI----------------GYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVV 405
              :|                  | |   | : | | |    |        |:       |   ||
Db        864 NGSSLSGTAEPGSTVILTDGNGNPIAEVTADGSGNWTYT-------PSTPIANGTVVNVV 916

Qy        406 QQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSA 465
               |  |                    | |:  :      :: ||||          :||
Db        917 AQDAAG-------------------NSSPGASVTVDSQAPAAPVVNP-----SNGTTLSG 952

Qy        466 HQPEEHIYDEVAADPGYSVIQNFSGSGPV-------TGRLIGTPGQGIQSTYALLANSGG 518
                        |:|| :|         |:       :|    |||  : :   : | :
Db        953 ----------TAEPGATVSLTDGNGNPIGQVTADGSGNWSFTPGTPLANGTVVNATASD 1001

Qy        519 LRLGMGGLTSGGETAVSSV-NAAPTPGP 545
                    |  ::   | | ||  |||   |
Db       1002 ---PTGNTSAPASTTVDSVAPAAPVVNP 1026


RESULT 10
US-09-328-352-4412
; Sequence 4412, Application US/09328352
; Patent No. 6562958
```

```
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4412
;   LENGTH: 734
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-4412

  Query Match             4.8%;  Score 134.5;  DB 4;  Length 734;
  Best Local Similarity   20.8%;  Pred. No. 0.0041;
  Matches  129;  Conservative   72;  Mismatches  202;  Indels  217;  Gaps   29;

Qy         74 SRLAAATSETCLLGGFEVLHDKGPLD--ILN----TQIGPSAFRVEVQADGTHAAIGEKN 127
              :|:  |: |       | :|| | | |: :||     ||| |      | |    |  ||
Db          8 TRIIVASKE-----GLDVLQD-GQLNKVVLNQPTIIQIGVS------QKD---IASMEKQ 52

Qy        128 GLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMV--TVASDIAEARTRILAKL 185
              | : : |   |   |::   |  |              : | :|  |        ||:
Db         53 GGSLVIHLKNGETIVLENFFNEATNT---------TEHSLVFPTEQGKFVEAQF------ 97

Qy        186 DPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVG-AIAAG 244
                   | | :| |  :        : :  |    :  |   |     :| :|| |  |:|  : ||
Db         98 ------DAQGKVIDYRGL-----NHVTDLAYTSTSPSAATMAVDNDPSF--SMGNVLKAG 144

Qy        245 LAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESAT-KDQLTQEAFKNPENQKVNI-D 302
              || ||| |:     |    :  | : : ||  | || :||  |         |    | |: | |
Db        145 LAVLAAEGL-YLWAFDKDKDDSPSTPDLIAPAAPTATLADDTVTVTGKTEANAKIYIKD 203

Qy        303 ANGNAIPSGE---------------LXDDIVEQIAQQAKEAGEVARQQAV---------- 337
              | || : ||                :  |  :  ||| |  ||  ::  |
Db        204 AAGNTVASGVADASGNYTIKLDKPLVNGDKLNVIAQDA--AGNNSKVTVVTGTKDTIAPD 261

Qy        338 -----------------ESNAQ----------------------------------AQQ 345
                               |:||:                                  :
Db        262 VPQAQLSDDGSLLTGKAEANAKITVYDATGKVLGTVFANKDGIYSLKLTPPLTSEAGGKV 321

Qy        346 RYEDQHARRQEELQLSSG------------------IGYGLSSA-----LIVAGG--IGA 380
                ||      : ||::: :                   : :|  :    |||   ||:
Db        322 VAEDAAGNKSEEVKIIAGKDTIPPASPFVEVNKEGSVIHGKTEANAKVQIKDAGGKVIGS 381

Qy        381 GVTTALHRRNQPAEQTTTTTT------HTVVQQQTGGIPQHKVALMP------------ 421
              |    |      |  | |       |||  :       | :  |
Db        382 GTADA-----QGEFQITLSPALKEAQKGTVVVEDAAGNVSKPVEITPGFDSIAPDKPTVQ 436

Qy        422 -----------QERRRFSDRRDSQGSVASTHWSDSSSEVV---------NPYAEVGGARN 461
                          |     : :|: |  |   :  :|:: :          | :| |      |
Db        437 INTDGTSVTGTAEANAKIEIKDTTGKVIGSGTADANGKFTISISPALTDNKHASVSAIDN 496

Qy        462 SLSAHQPEEHIYDEVAADPGYSVIQNFSGS-GPVTGRLIGTPGQGIQSTYALLANSG--- 517
               : : : : : :    |     :: :       | | |  :  | |         |  ||
Db        497 AGNKSEVVDIVGTKDTTPPAKPILNSVDDDVGAVKGAI--TAGSETDDARPKLTGSGEAN 554

Qy        518 --------GLRLGMGGLTSG 529
                      |: :|:  :|||
Db        555 ATLTIYDNGVAIGVVTVTSG 574
```

RESULT 11

```
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen I.
;    APPLICANT:  St. Geme III, Joseph W.
;    TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,995
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-409-995-4

  Query Match             4.8%;  Score 134.5;  DB 1;  Length 1912;
  Best Local Similarity   22.0%;  Pred. No. 0.018;
  Matches  124;  Conservative   62;  Mismatches  207;  Indels  171;  Gaps   26;

Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
              ||  ||                          | || :|: |       |   | |  |  |: :
Db       1162 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1197

Qy         66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
                      : | :       | |: : |  |   |     | ||               ||   ||
Db       1198 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1242

Qy        122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
              |: || ||                            |||      :      :|   ||:  |:|||:
Db       1243 AV--KNANEVEFV---------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1283

Qy        179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
                   |     || ||  |      ||  :  |   | |   : | ::   |:  ||:: |
Db       1284 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1342

Qy        230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
                          :| |   :|| :|   |   :       |||  :        | || | |     :
Db       1343 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1395

Qy        278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
              |  | |  | ||| :| |  | ||  |                 ||  :: |  : |  :
Db       1396 ANDALKAXDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1443

Qy        328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
               |     :  : |: :      :|        : | :||  |:  |      : :| || |
Db       1444 LGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1496

Qy        388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
                                    ||                    :    |::    |:    |
Db       1497 ----------------------GG----------------NGITDNEKKXAA-----SVK 1513

Qy        448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
              :|:|    | | : :  ||:    |:| | ||        |     :     ||      |  :  :    :
Db       1514 DVLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1568

Qy        508 STYALLANSGGLRLGMGGLTSGGE 531
               :   : | :  ::    | | :| |
```

```
Db       1569 TEVKIGAKTSVIKDHNGKLFTGKE 1592


RESULT 12
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/685,467
;      FILING DATE:  22-JUL-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053-2/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-685-467-4

  Query Match             4.8%;  Score 134.5;  DB 3;  Length 1912;
  Best Local Similarity   22.0%;  Pred. No. 0.018;
  Matches  124;  Conservative   62;  Mismatches  207;  Indels  171;  Gaps   26;

Qy           6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
               ||   ||                  |  || :|:  |       |   | |  |   |: :
Db        1162 LGGTANG-----------------RNDTGTVINKDG------LTITLANGAAAGTDASN 1197

Qy          66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL----DILNTQIGPSAFRVEVQADGTHA 121
                     : | :    |  |: : |  |    |  ||         | |    ||
Db        1198 --------GNTISVTKDGISAGNKEITNVKSALKTYKDTQNT-------ADETQDKEFHA 1242

Qy         122 AIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR--- 178
               |:  ||  ||                  |||    :     :|   ||:  |:|||:
Db        1243 AV--KNANEVEFV--------------GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1283

Qy         179 -----TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTST-TNSSVR--- 229
                  |   ||  ||  |    ||  : |   | |  : | ::  |:  ||:: |
Db        1284 GLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVTTDATTAQGTNANERGKV 1342

Qy         230 ------------SDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
                        :| |   :|| :|  :   ||| :        | ||  | |     :
Db        1343 VVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------KVENDDSATIDDSPTDDG 1395

Qy         278 AESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPSGELXDDIVEQIAQQAKE 327
                  |  | |  | || :| |   | ||  |         ||  :: |  : | :
Db        1396 ANDALKAXDTLTLKAGK---NLKVKRDGKNITFALANDLSVKSATVSDKL---------S 1443

Qy         328 AGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
                 |     :  : |: :     :|        : | :||    |:  |      : :| || |
Db        1444 LGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1496

Qy         388 RRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSS 447
                                    ||                  :   |::   |:     |
Db        1497 ----------------------GG----------------NGITDNEKKXAA-----SVK 1513

Qy         448 EVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQGIQ 507
               :|:|     | | : : ||:   |:| | ||    |  :   ||     |   : :   | :
Db        1514 DVLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSVTVESKDNGKR 1568

Qy         508 STYALLANSGGLRLGMGGLTSGGE 531
               :   : | :  ::   | | :| |
Db        1569 TEVKIGAKTSVIKDHNGKLFTGKE 1592


RESULT 13
US-09-252-991A-21021
; Sequence 21021, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 21021
;   LENGTH: 375
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21021

  Query Match             4.8%;  Score 134;  DB 4;  Length 375;
  Best Local Similarity   23.8%;  Pred. No. 0.0016;
  Matches   81;  Conservative   44;  Mismatches  158;  Indels   58;  Gaps   13;

Qy         195 PKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSD-PKFWVSVGAIAAGLAGLAATGI 253
               |:       :|:|         |  : |  : :    :| |:  :::   |   ||||  |
Db           6 PEAFQQVELGIGQGEAGAQLVPKDRHQGSADDVADADHPEVHLALHR-AVHPAGLATAGE 64

Qy         254 AQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIP---- 309
                 | |     ||      |:     |: | : : |     :: |      || | | |
Db          65 DDADAEENAADDRIRAHPEDLRRDAQ-ADRGEGTDAGGRHAEQ-----DAEGGAGPAGGP 118

Qy         310 -----SGELXDDIVEQIA-QQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQ-LSS 362
                  :||     ::| | |||::||     |: :|:    :|| | || : |  ||  :|
Db         119 EVPCGTGEADLRALDQAAEQQAEQAGHAQHQRYLEAPG-GEQRGERQHRQAQAGLQACAS 177

Qy         363 GIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQ--QQTGGIPQHKVALM 420
               |  :|         | |||      |  |  :|              :  |   | |
Db         178 GQAFG--------AGQGAGRGADFQRCAQERQQADLHADPGAADGGEHEGPGVQGKHRDT 229

Qy         421 PQERRRFSDRRDSQGSV-----------ASTHWSDSSSEVVNPYAEVGGA---------- 459
                :|||: :  | ::              |:     : ::|  |      ||
Db         230 GEERRQVAAHRQARAEAGDDTADQTLQQATFAGRPAQADVAGPEGAAQGAEEHADDHHSV 289

Qy         460 ----RNSLSAHQPEEHIYDEVAADPGYSVIQ---NFSGSGP 493
                   | :| | |||       : |:||  |       |:|  |
Db         290 DPVQRRALKADQPEVAPVLWLVAEPGQHVAAPAGEFAGHRP 330


RESULT 14
US-08-185-432-19
```

```
; Sequence 19, Application US/08185432
; Patent No. 5750652
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Busseau, Isabelle
;    APPLICANT:  Diederich, Robert J.
;    APPLICANT:  Xu, Tian
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  DELTEX PROTEINS, NUCLEIC ACIDS, AND
;    TITLE OF INVENTION:  ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/185,432
;      FILING DATE:  21-JAN-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2703 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-185-432-19

  Query Match             4.8%;  Score 133;  DB 1;  Length 2703;
  Best Local Similarity   19.8%;  Pred. No. 0.043;
  Matches  119;  Conservative   83;  Mismatches  193;  Indels  206;  Gaps   28;

Qy          4 GNLGNNVNGNHLIPPAPPLPSQT--DGAAR------GGTGHLISSTGALGSRSLFSPLRN 55
              || ||| |||     :    :||     ||:      :  : : :||:|  :   :  :
Db       2177 GNGGNNGNGN----ASGKQSNQTAKQKAAKKAKLIEGSPDNGLDATGSLRRK---ASSKK 2229

Qy         56 SMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQ 115
              : | |  : :: ||  || :|          ||   :    |    |: :  :|
Db       2230 TSAASKKAANLNGL--NPGQLT---------GGVSGVPGVPP---TNSAVQAAAAAAAAV 2275

Qy        116 ADGTHAAIGEKNGLEVSVTL-SPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDI 174
              |  :|   |   |: :   | || : ||: |
Db       2276 AAMSHELEGSPVGVGMGGNLPSPYDTSSMYS---------------------------- 2306

Qy        175 AEARTRILAKLDPDNHGGRQPKDVD-----TRSVGVGSASGIDDGVVSETHTSTTNSSVR 229
                |    ||  :| | | :||   :      :|:      :|: | :  : :  :  | :
Db       2307 -NAMAAPLANGNP-NTGAKQPPSYEDCIKNAQSMQSLQGNGL-DMIKLDNYAYSMGSPFQ 2363

Qy        230 SD-------------PKFWVSVGAIAAGLAGL-----AATG------------------ 252
              :              :  | | :  || |:     |  |
Db       2364 QELLNGQGLGMNGNGQRNGVGPGVLPGGLCGMGGLSGAGNGNSREQGLSPPYSNQSPPHS 2423

Qy        253 IAQALALTPE------------PDDPTTTDPDQAANAAESATKDQLTQEAF--------- 291
              :  :|||:|              |  ||:    ||   |          |: |
Db       2424 VQSSLALSPHAYLGSPSPAKSLPSLPTSPTHIQAMRHATQ-------QKQFGGSNLNSLL 2476

Qy        292 -------------------KNPENQKVN---IDANGN------AIPSGELXDDIVEQIA 322
                                : |:|  |:   |   |:      | |       |:: |:

Db       2477 GGANGGGVVGGGGGGGGGVGQGPQNSPVSLGIISPTGSDMGIMLAPPQSSKNSAIMQTIS 2536

Qy        323 QQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGV 382
              | ::      ||  :   | ||: : |  ::|:: |   |: :| |: | : |  |:
Db       2537 PQQQQ------QQQQQQQQQHQQQQQQQQQQQQQQQQQQLGGLEFG-SAGLDLNGFCGSPD 2589

Qy        383 TTALHRRNQPAEQT------------TTTTTH---------TVVQQQTGG-IPQHKVALM 420
              :   : | |: |:             : :: |         |   | :||  ||| |  :
Db       2590 SFHSGQMNPPSIQSSMSGSSPSTNMLSPSSQHNQQAFYQYLTPSSQHSGGHTPQHLVQTL 2649

Qy        421 PQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAE--VGGARNSL------SAHQPEEHI 472
                    |   :    : ||| ||    : ::|      | |:|        |::  | |
Db       2650 --------DSYPTPSPESPGHWSSSSPRSNSDWSEGVQSPAANNLYISGGHQANKGSEAI 2701

Qy        473 Y 473
              |
Db       2702 Y 2702


RESULT 15
US-08-899-232-4
; Sequence 4, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
;  APPLICANT: Artavanis-Tsakonas, Spyridon
;  APPLICANT: Qi, Huilin
;  TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
;  FILE REFERENCE: 7326-046
;  CURRENT APPLICATION NUMBER: US/08/899,232
;  CURRENT FILING DATE: 1997-07-23
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 2703
;   TYPE: PRT
;   ORGANISM: Drosophila sp.
US-08-899-232-4

  Query Match             4.8%;  Score 133;  DB 4;  Length 2703;
  Best Local Similarity   19.8%;  Pred. No. 0.043;
  Matches  119;  Conservative   83;  Mismatches  193;  Indels  206;  Gaps   28;

Qy          4 GNLGNNVNGNHLIPPAPPLPSQT--DGAAR------GGTGHLISSTGALGSRSLFSPLRN 55
              || ||| |||     :    :||     ||:      |  : : :||:|  :   :  :
Db       2177 GNGGNNGNGN----ASGKQSNQTAKQKAAKKAKLIEGSPDNGLDATGSLRRK---ASSKK 2229

Qy         56 SMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQ 115
              : | |  : :: ||  || :|          ||   :    |    |: :  :|
Db       2230 TSAASKKAANLNGL--NPGQLT---------GGVSGVPGVPP---TNSAVQAAAAAAAAV 2275

Qy        116 ADGTHAAIGEKNGLEVSVTL-SPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDI 174
              |  :|   |   |: :   | || : ||: |
Db       2276 AAMSHELEGSPVGVGMGGNLPSPYDTSSMYS---------------------------- 2306

Qy        175 AEARTRILAKLDPDNHGGRQPKDVD-----TRSVGVGSASGIDDGVVSETHTSTTNSSVR 229
                |    ||  :| | | :||   :      :|:      :|: | :  : :  :  | :
Db       2307 -NAMAAPLANGNP-NTGAKQPPSYEDCIKNAQSMQSLQGNGL-DMIKLDNYAYSMGSPFQ 2363

Qy        230 SD-------------PKFWVSVGAIAAGLAGL-----AATG------------------ 252
              :              :  | | :  || |:     |  |
Db       2364 QELLNGQGLGMNGNGQRNGVGPGVLPGGLCGMGGLSGAGNGNSREQGLSPPYSNQSPPHS 2423

Qy        253 IAQALALTPE------------PDDPTTTDPDQAANAAESATKDQLTQEAF--------- 291
              :  :|||:|              |  ||:    ||   |          |: |
Db       2424 VQSSLALSPHAYLGSPSPAKSLPSLPTSPTHIQAMRHATQ-------QKQFGGSNLNSLL 2476

Qy        292 -------------------KNPENQKVN---IDANGN------AIPSGELXDDIVEQIA 322
                                : |:|  |:   |   |:      | |       |:: |:

Db       2477 GGANGGGVVGGGGGGGGGVGQGPQNSPVSLGIISPTGSDMGIMLAPPQSSKNSAIMQTIS 2536

Qy        323 QQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGV 382
              | ::      ||  :   | ||: : |  ::|:: |   |: :| |: | : |  |:
Db       2537 PQQQQ------QQQQQQQQQHQQQQQQQQQQQQQQQQQQLGGLEFG-SAGLDLNGFCGSPD 2589
```

```
Qy        383 TTALHRRNQPAEQT-----------TTTTTH---------TVVQQQTGG-IPQHKVALM 420
              :    : | |: |:            : :: |         |   | :||  ||| |  :
Db       2590 SFHSGQMNPPSIQSSMSGSSPSTNMLSPSSQHNQQAFYQYLTPSSQHSGGHTPQHLVQTL 2649

Qy        421 PQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAE--VGGARNSL------SAHQPEEHI 472
                  |    :    :  ||| ||    : ::|     | |:|       |::  | |
Db       2650 --------DSYPTPSPESPGHWSSSSPRSNSDWSEGVQSPAANNLYISGGHQANKGSEAI 2701

Qy        473 Y 473
               |
Db       2702 Y 2702
```

Search completed: December 16, 2003, 08:57:53
Job time : 28.7653 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 53.0171 Seconds
(without alignments)
1643.636 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2798	99.9	549	20	AAY06220	EPEC E. coli trans
2	1492	53.3	559	20	AAY06221	EHEC E. coli trans
3	530	18.9	107	21	AAB20576	Intimin C-terminal
4	165	5.9	30	20	AAY06213	EPEC E. coli trans
5	151	5.4	934	22	ABB69655	Drosophila melanog
6	149	5.3	831	19	AAW68201	M. catarrhalis str
7	145.5	5.2	848	22	ABB61228	Drosophila melanog
8	145.5	5.2	2586	22	ABB66878	Drosophila melanog
9	144	5.1	883	22	ABB62230	Drosophila melanog
10	143.5	5.1	1046	22	ABB67028	Drosophila melanog
11	141.5	5.1	1488	22	AAG91741	C glutamicum prote
12	141	5.0	580	18	AAW10660	Endoglycoceramidas
13	140.5	5.0	1472	22	ABB59423	Drosophila melanog
14	139.5	5.0	985	23	ABP43746	2 PUMH2 protein.
15	139.5	5.0	2411	21	AAB23860	Haemophilus influe
16	138.5	4.9	2353	17	AAR99393	Haemophilus adhesi
17	138	4.9	1455	22	ABB63040	Drosophila melanog
18	138	4.9	2857	22	ABB64188	Drosophila melanog
19	136.5	4.9	3096	22	AAB46770	FANCIP3 protein fr
20	136.5	4.9	4630	18	AAW19629	Streptomyces venez
21	136	4.9	1064	22	ABB59068	Drosophila melanog
22	136	4.9	1569	22	AAG98842	E. coli growth and
23	136	4.9	2086	22	AAU34143	Staphylococcus aur
24	136	4.9	5795	22	AAU37017	Staphylococcus aur
25	135.5	4.8	1138	23	ABJ05434	Human breast cance
26	135	4.8	1045	22	ABG08332	Novel human diagno
27	135	4.8	1532	21	AAB40945	Human ORFX ORF709
28	135	4.8	1905	22	ABB64925	Drosophila melanog
29	135	4.8	2344	22	AAU37120	Staphylococcus aur
30	134	4.8	828	22	ABB58368	Drosophila melanog
31	133.5	4.8	620	22	ABB69938	Drosophila melanog
32	133.5	4.8	818	22	ABB70219	Drosophila melanog
33	133.5	4.8	1978	20	AAY27230	Amino acid sequenc
34	133	4.8	745	22	AAU25461	Human mddt protein
35	133	4.8	1596	22	ABB70845	Drosophila melanog
36	132.5	4.7	1026	18	AAW37490	Caulobacter cresce
37	132.5	4.7	1026	21	AAY44757	Caulobacter cresce
38	132.5	4.7	1026	24	AAE34374	Caulobacter cresce
39	132.5	4.7	1126	22	ABB60460	Drosophila melanog
40	132.5	4.7	4630	21	AAY77177	S. venezuelae vep
41	131.5	4.7	707	24	ABJ26513	Aspergillus fumiga
42	131.5	4.7	1026	15	AAR48993	rsaA S-lyaer prote
43	130.5	4.7	950	22	ABB62035	Drosophila melanog
44	130.5	4.7	1026	17	AAR94014	Caulobacter S-laye
45	130.5	4.7	2304	22	ABB60219	Drosophila melanog

ALIGNMENTS

RESULT 1
AAY06220
ID AAY06220 standard; Protein; 549 AA.
XX
AC AAY06220;
XX
DT 16-AUG-1999 (first entry)
XX
DE EPEC E. coli translocated intimin receptor (Tir).
XX
KW Tir; translocated intimin receptor; Hp90; enteropathogenic;
KW EPEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Domain 234..253
FT /note= "putative transmembrane domain"
FT Domain 364..386
FT /note= "putative transmembrane domain"
FT Misc-difference 180
FT /note= "encoded by AAA"
FT Misc-difference 314
FT /note= "given as Xaa in the specification; Lys
FT is deduced from the DNA sequence"
XX
PN WO9924576-A1.
XX
PD 20-MAY-1999.
XX
PF 10-NOV-1998; 98WO-CA01042.
XX

```
PR   12-NOV-1997;   97US-0065130.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Devinney R,  Finlay BB,  Kenny B,  Stein M;
XX
DR   WPI; 1999-337712/28.
DR   N-PSDB; AAX58858.
XX
PT   New translocated intimin receptor useful for treating infection by
PT   enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS   Claim 6; Page 55-58; 91pp; English.
XX
CC   The present sequence represents Tir, a novel translocated intimin
CC   receptor (formerly termed Hp90) from an enteropathogenic
CC   Escherichia coli (EPEC) strain.  The sequence was deduced from an
CC   isolated tir polynucleotide (see AAX58858).  Tir proteins are
CC   secreted by attaching and effacing pathogens such as EPEC and EHEC
CC   (see AAY06221) E. coli.  The bacterial pathogens insert their own
CC   receptors into mammalian cell surfaces, to which the pathogen then
CC   adheres to trigger additional host signaling events and actin
CC   nucleation.  Diagnosis of disease caused by pathogenic E. coli can
CC   be performed by use of antibodies that bind to Tir to detect the
CC   protein or the use of nucleic acid probes for detection of nucleic
CC   acids encoding Tir polypeptide.  Isolated Tir nucleic acids, Tir
CC   peptides, a recombinant method for producing recombinant Tir,
CC   antibodies which bind to Tir, and a kit for the detection of
CC   Tir-producing E. coli are provided.  A method of immunising a host
CC   with Tir to induce a protective immune response is also provided.
CC   In addition, Tir fusion proteins can be used in attenuated E. coli
CC   to induce a cell-mediated immune response to other polypeptides,
CC   e.g. antigens.  A method for screening for compounds which
CC   interfere with the binding of bacterial pathogens to their
CC   receptors is further provided.
XX
SQ   Sequence   549 AA;

  Query Match             99.9%;  Score 2798;  DB 20;  Length 549;
  Best Local Similarity   99.8%;  Pred. No. 2.6e-210;
  Matches  548;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60

Qy         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH 120

Qy        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEARTR 180

Qy        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGA 240

Qy        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVN 300

Qy        301 IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IDANGNAIPSGELKDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQL 360

Qy        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALM 420

Qy        421 PQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADP 480

Qy        481 GYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAA 540

Qy        541 PTPGPVRFV 549
              |||||||||
Db        541 PTPGPVRFV 549


RESULT 2
AAY06221
ID   AAY06221 standard; Protein; 559 AA.
XX
AC   AAY06221;
XX
DT   16-AUG-1999  (first entry)
XX
DE   EHEC E. coli translocated intimin receptor (Tir).
XX
KW   Tir; translocated intimin receptor; Hp90; enterohaemorrhagic;
KW   EHEC; infection; diagnosis; vaccine.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 453
FT                   /note= "encoded by codon of 1 apparent nucleotide,
FT                           causing frameshift in the DNA sequence"
XX
PN   WO9924576-A1.
XX
PD   20-MAY-1999.
XX
PF   10-NOV-1998;   98WO-CA01042.
XX
PR   12-NOV-1997;   97US-0065130.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Devinney R,  Finlay BB,  Kenny B,  Stein M;
XX
DR   WPI; 1999-337712/28.
DR   N-PSDB; AAX58859.
XX
PT   New translocated intimin receptor useful for treating infection by
PT   enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS   Claim 7; Page 55-58; 91pp; English.
XX
CC   The present sequence represents Tir, a novel translocated intimin
CC   receptor (formerly termed Hp90) from an enterohaemorrhagic
CC   Escherichia coli (EHEC) strain.  The sequence was deduced from an
CC   isolated tir polynucleotide (see AAX58859).  Tir proteins are
CC   secreted by attaching and effacing pathogens such as EHEC and EPEC
CC   (see AAY06220) E. coli.  The bacterial pathogens insert their own
CC   receptors into mammalian cell surfaces, to which the pathogen then
CC   adheres to trigger additional host signaling events and actin
CC   nucleation.  Diagnosis of disease caused by pathogenic E. coli can
CC   be performed by use of antibodies that bind to Tir to detect the
CC   protein or the use of nucleic acid probes for detection of nucleic
CC   acids encoding Tir polypeptide.  Isolated Tir nucleic acids, Tir
CC   peptides, a recombinant method for producing recombinant Tir,
CC   antibodies which bind to Tir, and a kit for the detection of
CC   Tir-producing E. coli are provided.  A method of immunising a host
CC   with Tir to induce a protective immune response is also provided.
CC   In addition, Tir fusion proteins can be used in attenuated E. coli
CC   to induce a cell-mediated immune response to other polypeptides,
CC   e.g. antigens.  A method for screening for compounds which
CC   interfere with the binding of bacterial pathogens to their
CC   receptors is further provided.
XX
SQ   Sequence   559 AA;
```

```
  Query Match             53.3%;  Score 1492;  DB 20;  Length 559;
  Best Local Similarity   54.7%;  Pred. No. 4.2e-108;
  Matches  322;  Conservative   64;  Mismatches  133;  Indels   70;  Gaps   12;

Qy          1 MPIGNLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
              |||||||:| | |: ||||||||||||||  || | ||:||| ||||:||:|:|||||||
Db          1 MPIGNLGHNPNVNNSIPPAPPLPSQTDGA--GGRGQLINSTGPLGSRALFTPVRNSMADS 58

Qy         61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG 118
               |:|  |:|||| || |||| ||  |  ||||||| |||| || ||| | |||| | ||
Db         59 GDNRASDVPGLPVNPMRLAA--SEITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDG 116

Qy        119 THAAIGEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
               | |:|::||:| || || ||:: ||||| |||::|||||||||:|| ||||||| |||
Db        117 KHIAVGQRNGVETSVVLSDQEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEAR 176

Qy        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGI------DDGV--VSETHTSTTNSSVRS 230
               |||  |:|  |           | | : |:      : |    :|| |||: ||:||
Db        177 QRILELLEPKGTG---------ESKGAGESKGVGELRESNSGAENTTETQTSTSTSSLRS 227

Qy        231 DPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEA 290
               ||| |:::| :| || ||||||| |||||||||| ||||||| ||:| |:||:||||:||
Db        228 DPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKEA 287

Qy        291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQ 350
               |:||:|||||||  ||||||| | ||:|  | :||| ||| |:|||:|:||||::|::|
Db        288 FQNPDNQKVNIDELGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAIENNAQAQKKYDEQ 347

Qy        351 HARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT-------HT 403
               |:||||::||| ||||| |||: ||||  || ||||:||| ||||||||        |
Db        348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVEQTTTTTTTTTTSART 407

Qy        404 VVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGG---AR 460
               |  :     |       |         || |  | :|| :  ||          :||    |
Db        408 VENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSSTFFDTSS---------IGGPCRIR 458

Qy        461 NSLSAHQ--------------------PEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIG 500
                 :  |:                          :|  :   |     :: : :|    ||:|
Db        459 MLMLKHRCMIRRCRLLILIRLFRIWGIQISVVYSTIQHPP-----RDTTDNG---ARLLG 510

Qy        501 TPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAAPTPGPVRFV 549
                |  ||||||| || |||||  ||||| |  :||:: |  | ||  |||
Db        511 NPSAGIQSTYARLALSGGLRHDMGGLTGGSNSAVNTSNNPPAPGSHRFV 559


RESULT 3
AAB20576
ID   AAB20576 standard; protein; 107 AA.
XX
AC   AAB20576;
XX
DT   08-DEC-2000  (first entry)
XX
DE   Intimin C-terminal Tir binding domain amino acid sequence.
XX
KW   Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
KW   Tir-independent eukaryotic cell binding activity; bacterial infection;
KW   diarrhoea; antibacterial.
XX
OS   Unidentified.
XX
PN   WO200045173-A1.
XX
PD   03-AUG-2000.
XX
PF   31-JAN-2000; 2000WO-GB00254.
XX
PR   29-JAN-1999;   99GB-0001897.
XX
PA   (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI   Frankel GM,  Matthews SJ,  Hale CB,  Dougan G;
XX
DR   WPI; 2000-499357/44.
XX
PT   Screening for inhibitors of intimin binding to eukaryotic cells, for
PT   use in diagnosing, preventing and treating bacterial infections,
PT   especially Escherichia coli O157:H7  -
XX
PS   Claim 8; Page 76; 96pp; English.
XX
CC   The present invention describes a method of screening for an inhibitor
CC   of intimin binding to eukaryotic cells. The method comprises exposing an
CC   intimin polypeptide having a Tir-independent cell binding activity to
CC   test agents, and obtaining an inhibitor based on its ability to bind the
CC   polypeptide. The inhibitors are used in the prevention, treatment and/or
CC   diagnosis of bacterial infections, preferably by enteropathic and/or
CC   enterohaemorrhagic Escherichia coli, Shiga toxigenic E. coli, Hafnia
CC   alvei or Citrobacter freundii, or especially E. coli O157:H7. The
CC   infections cause a histopathological effect known as attachment and
CC   effacement on intestinal epithelial cells. The inhibitors can be used
CC   to produce food supplements or additives, especially where the food is
CC   a milk substitute. The method can be used to sort cells based on their
CC   ability to bind to a Tir independent cell binding domain of an intimin
CC   polypeptide. Polypeptides having Tir-independent intimin binding
CC   activity can be used to produce a vaccine against a bacterial disease.
CC   The present sequence represents a specifically claimed intimin C-terminal
CC   Tir binding domain amino acid sequence, for use in the method of the
CC   present invention.
XX
SQ   Sequence   107 AA;

  Query Match             18.9%;  Score 530;  DB 21;  Length 107;
  Best Local Similarity   99.1%;  Pred. No. 8.3e-34;
  Matches  106;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        255 QALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIPSGELX 314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIPSGELK 60

Qy        315 DDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLS 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLS 107


RESULT 4
AAY06213
ID   AAY06213 standard; Peptide; 30 AA.
XX
AC   AAY06213;
XX
DT   16-AUG-1999  (first entry)
XX
DE   EPEC E. coli translocated intimin receptor N-terminal peptide.
XX
KW   Tir; translocated intimin receptor; Hp90; enteropathogenic;
KW   EPEC; infection; diagnosis; vaccine.
XX
OS   Escherichia coli.
XX
PN   WO9924576-A1.
XX
PD   20-MAY-1999.
XX
PF   10-NOV-1998;   98WO-CA01042.
XX
PR   12-NOV-1997;   97US-0065130.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Devinney R,  Finlay BB,  Kenny B,  Stein M;
XX
DR   WPI; 1999-337712/28.
XX
PT   New translocated intimin receptor useful for treating infection by
```

```
PT   enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS   Example 1; Page 37; 91pp; English.
XX
CC   The present sequence represents the N-terminal sequence of Tir (see
CC   also AAY06220), a novel translocated intimin receptor from an
CC   enteropathogenic Escherichia coli (EPEC) strain. The 78 kDa EPEC
CC   protein is secreted by the bacterial pathogen. Diagnosis of
CC   disease caused by pathogenic E. coli can be performed by use of
CC   antibodies that bind to Tir to detect the protein, or the use of
CC   nucleic acid probes for detection of nucleic acids encoding Tir. A
CC   kit for the detection of Tir-producing E. coli is provided. Also
CC   provided are a method of immunising a host with Tir to induce a
CC   protective immune response, and a method for screening for
CC   compounds which interfere with the binding of bacterial pathogens
CC   to their receptors.
XX
SQ   Sequence   30 AA;

  Query Match             5.9%;  Score 165;  DB 20;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-06;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 PIGNLGNNVNGNHLIPPAPPLPSQTDGAAR 31
              ||||||||||||||||||||||||||||||
Db          1 PIGNLGNNVNGNHLIPPAPPLPSQTDGAAR 30


RESULT 5
ABB69655
ID   ABB69655 standard; Protein; 934 AA.
XX
AC   ABB69655;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 35757.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13758.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 35757; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   934 AA;

  Query Match             5.4%;  Score 151;  DB 22;  Length 934;
  Best Local Similarity   19.6%;  Pred. No. 0.0074;
  Matches  111;  Conservative   74;  Mismatches  218;  Indels  164;  Gaps   21;

Qy         18 PAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLA 77
              ||     :  :|:|  |:    | :| :         ||:|  : :|         |  : :
Db        322 PAEDTTANRNGSAGSGSS---SKSGVIPKNR--KSLRHSQQNLLD--------TEAAEVE 368

Qy         78 AATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTH---------------- 120
              :| :||   | |   : | ||:::    | |   : :| | |
Db        369 SAAAET----GGEQQSEHGNLDVIDQLPTPEAGPNQSEASGIHNQSQLEADRQPVVQDED 424

Qy        121 ---------------------AAIGEKNGLEVSVTLSPQEWSSLQSIDTE--------- 149
                                  : |     :  :     | :: : :| |
Db        425 DEEDDEEEEEEEEEEVGFYGIVNSAGSSYEEDTQIVAEEDEITTEEEVDDEDDDEIEEED 484

Qy        150 -GKNRFV--FTGGRGG---------------SGHPMVTVASDIAEARTRILAKLDPDNH 190
               :: |     |  ||                :  | |  || |    |   :
Db        485 LSESEFAQQLIGELGGERQQPQRATNNAKKTKNNTPAVAAASTIKATTTATTS------- 537

Qy        191 GGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAA 250
              | | :     | |:| |   |    | :| :|||        |   :|   |  :|
Db        538 SGYQQR---AAPHGQGAAPGAGGGGPRATRSSNSNSSNN------VVDYSIMPHLTTVAT 588

Qy        251 TGIAQALALTPEPDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAIPS 310
              |      ||||:        |   ::  :   : |   :    | :|  ::  :    |:
Db        589 T---TPHALTPQQHQQQRQPPQPPSHQQQQVHQQQQPPQLL--PTHQFAHL--SSFVTPT 641

Qy        311 GELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSG--IGYGL 368
                    |  |  |       |  :    || :: ||  :|: :| :      |
Db        642 A---------AAAAAAAAAHYPPASAAAAYFAQQQHHQHQHQHQQQHVQQAQAPQPHYYH 692

Qy        369 SSALIV-------AGGIGAGVTTALHRRNQPAEQT-----TTTTTHTVVQQQTGGIPQHK 416
              ||  ::           |||   || ::  | : |     | ||  | | |     | :
Db        693 SSVAVLHQPPPHHFTSTGAGPPPALFQQQAPPQLTRYTPATATTAATFVPPQ-----QQQ 747

Qy        417 VALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEV 476
              |   ||:::  | || |:|                  : |   :| :| | ::|   :
Db        748 VVYNPQQQQHSSLRRSSRG------------------KTGSCVSSAAAAQQQQHHQQQ- 787

Qy        477 AADPGYSVIQNFSGSGPVTGRLIGTPG 503
                    |: | :  |  :|:  ||
Db        788 ---------QHSSAAAAVQQQLLPPPG 805


RESULT 6
AAW68201
ID   AAW68201 standard; Protein; 831 AA.
XX
AC   AAW68201;
XX
DT   07-OCT-1998  (first entry)
XX
DE   M. catarrhalis strain O35E UspA1 antigen.
XX
KW   Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
KW   vaccine; otitis media; sinusitis; lower respiratory tract infection;
KW   immunity enhancer; immunoassay reagent.
XX
OS   Moraxella catarrhalis.
XX
PN   WO9828333-A2.
XX
PD   02-JUL-1998.
XX
PF   19-DEC-1997;   97WO-US23930.
XX
```

```
PR   20-DEC-1996;   96US-0033598.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Aebi C,  Cope LD,  Fiske MJ,  Fredenburg R,  Hansen EJ;
PI   Maciver I;
XX
DR   WPI; 1998-377595/32.
DR   N-PSDB; AAV41341.
XX
PT   New peptide(s) containing the core epitope of Moraxella catarrhalis
PT   Usp proteins - useful in, e.g. vaccines to prevent or treat M.
PT   catarrhalis infection, and antibodies for passive immunisation
XX
PS   Claim 22; Pages 133-136; 237pp; English.
XX
CC   This represents a UspA1 antigen of Moraxella catarrhalis strain O35E.
CC   Nucleic acid sequences encoding the UspA1 and A2 antigens of
CC   M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in
CC   genetic vaccination. An antigenic composition or vaccine containing
CC   antigenic peptides from UspA1 or UspA2 antigens are used to induce an
CC   immune response in mammals against M. catarrhalis and can be used to
CC   treat infections such as otitis media, sinusitis, lower respiratory
CC   tract infections. They can also be used as immunity enhancers for other
CC   bacterial, parasitic or viral antigens, to raise antibodies and as
CC   immunoassay reagents for detecting specific antibodies. The antibodies
CC   are useful for passive immunisation and as immunoassay reagents.
CC   Detection of the epitopic core sequence, by immunoassay or by PCR, is
CC   used to diagnose infection. The Usp antigens encoding nucleic acid
CC   sequences are also used to produce recombinant proteins and for screening
CC   for potential anti-M. catarrhalis agents, while their fragments are
CC   useful as diagnostic probes or primers or to isolate variant sequences.
XX
SQ   Sequence   831 AA;

  Query Match             5.3%;  Score 149;  DB 19;  Length 831;
  Best Local Similarity   19.2%;  Pred. No. 0.0091;
  Matches  101;  Conservative   73;  Mismatches  193;  Indels  160;  Gaps   21;

Qy         74 SRLAAATSETCLLGGFEVLHDKGPLDILNTQIGP---SAFRVEVQADGTHAAIGEKNGLE 130
              |     || : |||       |    : | ::|    : |  | :|:    ::| : |
Db        226 SGTTGVTSNSVLLGNETA--GKQATTVKNAEVGGLSLTGFAGESKAENGVVSVGSEGGER 283

Qy        131 VSVTLSPQEWSSLQSIDTEGKNRFVFTGGRGGSGHPMVTVAS----DIAEARTRILAKLD 186
                | :       :    |: |         |   | : ||      ||   |  ||   |
Db        284 QIVNVGA---GQISDTSTDAVN--------GSQLHALATVVDDNQYDIVNNRADILNNQD 332

Qy        187 P-----------DNHGGRQPKDVDT----------------------RSVGVGS----- 207
                        ||  |   :|:::                        ||| |
Db        333 DIKDLQKEVKGLDNEVGELSRDINSLHDVTDNQQDDIKELKRGVKELDNEVGVLSRDINS 392

Qy        208 -----ASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPE 262
                  |   ||   ::       |  |:   |    || ::  :  |
Db        393 LHDDVADNQDDIAKNKADIKGLNKEVKELDK---EVGVLSRDIGSL-------------- 435

Qy        263 PDDPTTTDPDQAANAAESATKDQLTQEAFKNPE----NQKVNIDAN-------------- 304
                ||  |   | | | |:  | :   :|    |       :|| :|| |
Db        436 HDDVATNQADIAKNQADIKTLENNVEEELLNLSGRLLDQKADIDNNINNIYELAQQQDQH 495

Qy        305 -------GNAIPSG--ELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQ 355
                    | :  |  :|   :::| |  ||   ::|: |    :  |    :||:|::|
Db        496 SSDIKTLKNNVEEGLLDLSGRLIDQKADIAKNQADIAQNQTDIQDLAAYNELQDQYAQKQ 555

Qy        356 EE-------------------LQLSSGIG--YGL-------SSALIVAGGIGAGVTTAL 386
                |                  ::: |   | |        || :    : |  |  :
Db        556 TEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKVSAANTDRI 615

Qy        387 HRRNQPAE---QTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWS 443
                :    |:   :| |   :|:::|    : |:|   : ||   |:   | |
Db        616 AKNKAEADASFETLTKNQNTLIEQGEALVEQNKA--INQELEGFAAHADIQ--------- 664

Qy        444 DSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSG 490
                  :::   |::   : ::     |::|   ||   |: : :| :|
```

```
Db        665 --DKQILQNQADITTNKTAI-----EQNINRTVA--NGFEIEKNKAG 702


RESULT 7
ABB61228
ID   ABB61228 standard; Protein; 848 AA.
XX
AC   ABB61228;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10476.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05331.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10476; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   848 AA;

  Query Match             5.2%;  Score 145.5;  DB 22;  Length 848;
  Best Local Similarity   20.7%;  Pred. No. 0.017;
  Matches  143;  Conservative   79;  Mismatches  241;  Indels  229;  Gaps   30;

Qy          6 LGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
              ||  | |  ::  |   | |       | |:  ||  |        ||  :|  | ||
Db         39 LGVGV-GVGILSDASLSPIQQGSGGHSGGGNTNSSPLAPNG----VPLLTTMHRSPDSPQ 93

Qy         66 IPGLP--TNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADG----- 118
                | |   || : |   |  :        |:||  : |  |        :| : |||
Db         94 -PELATMTNVNVLDLHTDNS-------KLYDKEAVFIYETP------KVVMPADGGGGNN 139

Qy        119 ------THAAIGEKNGLEVSVTLSPQ-----------EWSSLQSIDTEGKN--------R 153
                   | |  : | :        |             |:   | |   | |         |
Db        140 SDEGHAIDARIAAQMGNQAQQQQQQQQQTEHQPLAKIEFDENQIIRVVGPNGEQQQIISR 199

Qy        154 FVFTG-----GRGGSG-HPMVTVASD----------IAEARTRILAKLDPDNHGGRQPKD 197
                :  |       |  :| | :  : ||          :| |       |:: | ||
Db        200 EIINGEHHILSRNEAGEHILTRIVSDPSKLMPNDNAVATAMYNQAQKMNND-HG------ 252
```

```
Qy        198 VDTRSVGVGSASGIDDGVV-----SETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAAT- 251
Db        253 ---QAVYQTSPLPLDASVLHYSGGNDSNVIKTEADIYEDHKKHAAAAAAAAGGGSIIYTT 309

Qy        252 ------GIAQALALT-PEPDDPTTTDPDQ--------AANAAESATKDQLTQEAFKNPE- 295
Db        310 SDPNGVNVKQLPHLTVPQKLDPDLYQADKHIDLIYNDGSKTVIYSTTDQKSLEIYSGGDI 369

Qy        296 ----------------------NQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEV 331
Db        370 GSLVSDGQVVVQAGLPYATTTGAGGQPVYIVADG-ALPAG------VEEHLQSGKLNG-- 420

Qy        332 ARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGL-----SSALIVAGGIGAGVTTAL 386
Db        421 ------------QTTPIDVSGLSQNEIQ-----GFLLGSHPSSSATVSTTGVVSTTTISH 463

Qy        387 HRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSS 446
Db        464 HQQQQQQQQ----------QQQ------------QQQQQQHQQQQQHPGDIVSAAGVGST 501

Qy        447 SEVVNPYAEVGGARNSLS-AHQPEEHIYD---------EVAADPGYSVIQNF-------- 488
Db        502 GSIVSSAAQQQQQQQLISIKREPEDLRKDPKNGNIAGAATANGPGSVITQKILHVDAPTA 561

Qy        489 --------------------SGSGPVTGRLIGTPGQGIQSTYALLANSGGLRL------- 521
Db        562 SEADRPSTPSSSINSTENTESDSQSVSGSESGSPGARTTATLEMYATTGGTQIYLQTSHP 621

Qy        522 --------GMGGLTSGGETAVSSVNAAPTPGP 545
Db        622 STASGAGGGAGPAGAAGGGGVSMQAQSPSPGP 653
```

RESULT 8
ABB66878
ID ABB66878 standard; Protein; 2586 AA.
XX
AC ABB66878;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 27426.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL10981.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 27426; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2586 AA;

Query Match 5.2%; Score 145.5; DB 22; Length 2586;
Best Local Similarity 21.0%; Pred. No. 0.081;
Matches 124; Conservative 64; Mismatches 273; Indels 129; Gaps 23;

```
Qy          5 NLGNNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVD-- 62
Db         78 NYVNCVNGKKQLPLLDG-SSITDHSASTNTITKIIDDG------LSSQTTSSSAPVVDVT 130

Qy         63 ---SRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGT 119
Db        131 QGSSSNGDGNSTQSSTTTTTTTTTSSDGG-EFTTSSDPV------------VEVSQGTN 176

Qy        120 HAAIGEKNGLEVSVTLSPQEWSSLQSID-----TEGKNRFVFTGGRGGSGHPMVTVASDI 174
Db        177 GGNSSTQSSSSTTTTTSSDEGQTTSSSDPVVEVAQGSS----SNGDGNSTQSLTTTTTTT 232

Qy        175 AEARTRILAKLDPDNHGGRQPKDVD-TRSVGVGSASGIDDGVVSETHTSTTNS-----SV 228
Db        233 TTS-----------SDGGQSTTSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQTTS 281

Qy        229 RSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPD---DPTTTDP----DQAANAAESA 281
Db        282 SSDPVVEVAQGSSSNGDGNSTQSSTTTTTTTTTSSDGGQSTTLSDPVVEVSQGTNGGNSS 341

Qy        282 TK-DQLTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQAVESN 340
Db        342 TQSSSSTTTTTSSDEGQTTS------------SSDPVGEVAQGSSSNGDGNSTQSSTTT 388

Qy        341 AQAQQRYED---QHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTT 397
Db        389 TTTTTTSSDGGQSTTSSDPVVEVSQGTNGGNSS------------------TQSSSATT 429

Qy        398 TTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSE------VVN 451
Db        430 TTTSSDEGQTTSSSDPVSEVA---QGSSSIGDGNSTQSSTTTTTTTTTSSDGGQSTTSSD 486

Qy        452 PYAEVG-GARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQG----- 505
Db        487 PVVEVSQGTNGGNSSTQSSSSTTTTTSSDEG----QTTSSSDPVVEVAQGSSSNGDGNST 542

Qy        506 ----IQSTYALLANSGG---------LRLGMGGLTSGGETAVSSVNAAPT 542
Db        543 QSSTTTTTTTTTSSDGGQSTTSSDPVVEASQG--TNGGNSSTQSSSSTTT 590
```

RESULT 9
ABB62230
ID ABB62230 standard; Protein; 883 AA.
XX
AC ABB62230;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13482.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

```
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06333.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 13482; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   883 AA;

  Query Match             5.1%;  Score 144;  DB 22;  Length 883;
  Best Local Similarity   20.8%;  Pred. No. 0.024;
  Matches  124;  Conservative   68;  Mismatches  197;  Indels  208;  Gaps   26;

Qy         39 SSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHD---- 94
              ||: :| | |  | | ::    | |   |  ||:|   ||| :|     | |:  |
Db         38 SSSLSLSSSSSSSSLSSATPTPVAS---PVTPTSPPPAAAAPAEASPPAGAELQEDGQQA 94

Qy         95 ---KGPL----DILNTQIGPSAFRVEVQADG----THAAIGEKNGLEVSVTLSPQEWSSL 143
                 : |     |:|           | |:  |    | : :  ::   |: |  |:   ||
Db         95 KTQEDPTMKDQDMLEKTRQEVKDPVNVEEPGAIVDTESVMARQSPSPVASTKVPE---SL 151

Qy        144 QSIDTEGKNRFVFTGGRGGSGHPMV--------TVASDIAEARTRILAKLDPDNHGGRQP 195
              : | :                | |        :|||| | :         :|    |
Db        152 EEISNKS--------------PPVQEDEEESESVASDCREFKVLYNHLRQQQHH--HSP 194

Qy        196 KDVD-TRSVGVGSASGIDD--GVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATG 252
              | |||        :||   :: |       |||
Db        195 SSPDKTRST-------LDDVSKILWERKQQLQRSSV----------------------- 223

Qy        253 IAQALALTPE---------PDDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDA 303
              |  |  | |:          |: |  | | |     |:  :|:   | ::|  :  :::
Db        224 ITAAPTLQPQQHQQPMSDIEDEETLEDVDDADADVEADAEDEELLEQYQNGYDSPLDLSL 283

Qy        304 NG------------NAIP-------SGELXDDIV--EQIAQQAKEAGEVARQQAVE---- 338
              |            :|:         ||   ||    |:   : |   :  |  | :
Db        284 GGATSAAASAAAAASAVSRRRGRTYSGTESDDSAQCERARMRLKPERKAERSAAYKKSLM 343

Qy        339 --------------SNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTT 384
                           | |  ||  :  | ::|:: |  :|  :              || |
Db        344 KRYYTEIPIVKQSTSPAPQQQLQQQHHLQQQQQQQPHNGSTF------------AGATA 390

Qy        385 ALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSD 444
              || :   || |  |   : |||                 :  ||
Db        391 LLHIK---TEQNTLLTPLQLQQQQ---------------------QQQQG--------- 416

Qy        445 SSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPV---------- 494
                  ::  |  ||: |  :||| :: :     | |   :: |    | :
Db        417 -----LHGAAGNGGSSNGNNAHQQQQPL-----AIPQRPLLHNLLSGGAIHNPHHRNYTT 466

Qy        495 --TGRLIGTPG-QGIQSTYALLANSGG--------LRLGMGGLTSGGETAVSSVNAA 540
                 ||    :|   |:    :  ::|||          ||||   ||    | ::  ||
Db        467 ATTGSFPPSPADSGVSDVDS--SSSGGQPCADELKARLGMPPATSASAAAAAAAAAA 521


RESULT 10
ABB67028
ID   ABB67028 standard; Protein; 1046 AA.
XX
AC   ABB67028;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 27876.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11131.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 27876; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1046 AA;

  Query Match             5.1%;  Score 143.5;  DB 22;  Length 1046;
  Best Local Similarity   19.9%;  Pred. No. 0.033;
  Matches  120;  Conservative   66;  Mismatches  245;  Indels  173;  Gaps   24;

Qy         28 GAARGG---TGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNP--SRLAAATSE 82
              | : |    :| |:||: :  | ||    : |   |        ||   |  |||  : |
Db        210 GVSSGSVSLSGQLLSSSASGSSGSLAGGQQASQTPS-------GLQPTPRKSRLKRSKSP 262

Qy         83 TCLLGGF----------------EVLHDKGPLDILNTQIGP-SAFRVEVQADGTHAAI-- 123
                 ||                :  |   |  ||     | |   ::  ||| :  :
Db        263 DLSSGGGAGSSGGSSSGSTQQQPQPAHHHHPQTILIQGQNPNSIVSLQQTADGNYIPVSG 322

Qy        124 -GEKNGLEVSVTLSPQEWSSLQSIDTEGKNRFVFTGGRG----GSGHPMVTVASDIAEAR 178
               |: :  |                | | ::    | |     | ||           | :
Db        323 SGDDSDAE---------------DHEHEHNHGHGHGHGGHTHGHGHSHEHSHDHEHEHK 366
```

```
Qy        179 TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSV 238
Db        367 PNKICKTE---HSVASP--ASNSSSASNNAAGV-SGVTSTGQAIVTQIVVARDGKDTKNM 420

Qy        239 GAIAAGLAGLAATGIAQA-LALTPEPDDPT----------------TTDPDQAANAAESA 281
Db        421 TSLGMGMNG-GLLGVPMGFLDFTPEPPAPSATPVTVTEHVDLSCNPSTDTRDLSNTTEPL 479

Qy        282 TKD--------QLTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVAR 333
Db        480 DIDNHLAQQIHRLDQSPMHSISHHHTGDESNSN----------LVQHIKSEVIEAKHLAA 529

Qy        334 QQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPA 393
Db        530 QHHALSQAQQQHAHHQAHQQHQQHQQQQ-----------------------HQQQQQH 564

Qy        394 EQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVASTHWSDSSSEVVNPY 453
Db        565 LHAQQLLAQSQLQQQ-----QQQQQQQQQHHQQQQQQAAAAAAAAGVHGQHGGHVT---H 616

Qy        454 AEVGGAR-----NSLSAHQPEEHIYDEVA----------------------------A 478
Db        617 ADIGGATVMEIDPSQIKHEPGMIITPEIVNMMSSGHMDMYNSDTSEDSMMIANGSPHDQK 676

Qy        479 DPGYSVIQNFSGSGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVN 538
Db        677 EPHYTNLDQQHGLG---GSVCG-PGPG-------GAGGGGGMSGGAGSGSGEKDALKNQN 725

Qy        539 AAPT 542
Db        726 MSLT 729
```

RESULT 11
AAG91741
ID AAG91741 standard; Protein; 1488 AA.
XX
AC AAG91741;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5495.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH66960.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 5495; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1488 AA;

Query Match 5.1%; Score 141.5; DB 22; Length 1488;
Best Local Similarity 18.9%; Pred. No. 0.078;
Matches 106; Conservative 75; Mismatches 225; Indels 155; Gaps 19;

```
Qy         48 SLF-SPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIG 106
Db        778 ALFEQPLADSLSSNESAELVAGAGMGLVGMALATSGNWAIAGMVI--------------- 822

Qy        107 PSAFRVEVQADGTHAAIGEKNGLEVSVTLSPQEWSSLQS-ID---TEGKNRFVFT----- 157
Db        823 --------------TVASSMGVIIFTIIAMKVRGSLVSGVDEAVTSVINRFLDTQVSSA 867

Qy        158 GGRGGSGHPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVS 217
Db        868 GATSGDGMMRRAAATGLGIGATHMVLNRDGDGGG----SDSGSGGSGGGSDSGLGEKAAG 923

Qy        218 ETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPDQAANA 277
Db        924 LAKVVTT---------------VAGAGLVGKYATD-----ALDNYADGVINGDGDGAFAA 963

Qy        278 AESATKDQLTQEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAKEA--------- 328
Db        964 GGDATVDG--------------DYVADGDAIASADANADFVDGVVDGAGRASFSNAAYSS 1009

Qy        329 -GEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGI-GAGVTTAL 386
Db       1010 DGTTLDGEGASVDAQGNPLHADGTPMSAAEAEMKMA-GLSSSGTMMEKSGVKSSGITTAA 1068

Qy        387 HRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDRRDSQGSVAST------ 440
Db       1069 DVMDDQSLASSVTESGLSKIPDTYG---------------ADVSGAAGTVGTTGADYSA 1112

Qy        441 -------HWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGS-- 491
Db       1113 TDSSAGLNMSEAALQSGTPMGALAGGSVS-SSDQAMNDAALQIAASQGLAPAGSIAGMEQ 1171

Qy        492 ----------------GPVTGRLIGTP--------GQGIQSTYALLANSGGLRLGMGGL 526
Db       1172 LSAQATEAPAGKAGKQLGDLSGSALNTQLASMGQQVGDSVNSAYA----AG----GMGGV 1223

Qy        527 TSGGETAVSSVNAAPTPGPVR 547
Db       1224 DVAGKVTEAAQHLSQVPGQIQ 1244
```

RESULT 12
AAW10660
ID AAW10660 standard; Protein; 580 AA.
XX
AC AAW10660;
XX
DT 20-JUL-1997 (first entry)
XX
DE Endoglycoceramidase activator II.
XX
KW Endoglycoceramidase; activator; glycolipid; enzyme.

XX
OS Rhodococcus spp. M-777.
XX
PN EP759470-A2.
XX
PD 26-FEB-1997.
XX
PF 28-JUN-1996; 96EP-0110513.
XX
PR 29-JUN-1995; 95JP-0188466.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Ito M, Izu H, Izumi Y, Kato I, Kurome Y, Sano M;
XX
DR WPI; 1997-147519/14.
DR N-PSDB; AAT61050.
XX
PT DNA encoding endo:glyco:cer:amidase activator polypeptide - for
PT prodn. of recombinant polypeptide, useful in glyco:lipid analysis
XX
PS Claim 1; Page 13-15; 30pp; English.
XX
CC DNA encoding endoglycoceramidase activator polypeptide can be used
CC for the prodn. of the recombinant polypeptide, which can be used
CC to research the intracellular function of glycolipids, because
CC it shifts the optimum pH for endoglycoceramidase II towards
CC neutral, allowing the latter to be used to hydrolyse glycolipids
CC even at pH 7.5.
XX
SQ Sequence 580 AA;

Query Match 5.0%; Score 141; DB 18; Length 580;
Best Local Similarity 23.9%; Pred. No. 0.023;
Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;

```
Qy          4 GNLGNNVNGNHLIPP----------APPLPSQTDGAARGGTGHLISSTGALGSRSLFSPL 53
              |  ||: |    :|           ||   |  | ||    | | :  | | :  :   |
Db        215 GQFGNDANFLTFLPRASAPIVGTVWAPTQCSPRDTAA----GPLNAGAGPLATIQI---L 267

Qy         54 RNSMADSVDSRDIPGLPTNPSR--LAAATSETCLLGGFEVLHD----KGPLDILNTQIGP 107
              | ::| :|   | |   ||     | |     |   |  :    |       |:|::|   |
Db        268 RQAVA-TVSYLDGPSAVTNGGEFTLNATVVPTPDSGQVQFTRDGEDVGAPVDLVN---GK 323

Qy        108 SAFRVEVQADGTHAAIGEKNGLEV--------SVTLSPQEWSSLQSIDTEGKNRFVFTGG 159
              ::    :  || :|   :  | |         :||:: |:  :  |:
Db        324 ASLTQSLDTDGDYAYEAKFLGAEFFNPSSAAKTVTVTSQDIQTTTSVT------------ 371

Qy        160 RGGSGHPMVTVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSET 219
                |  |          :    : ||::|   ||    :||      ||:|  :|||
Db        372 --GPDH------DAYRDQPVNLTAKVEPGVSGGTVAFEVD--GTPVGTADVMDDGAAVLP 421

Qy        220 HTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALAL--------TPEPDDPTTTDP 271
              || ||| : |           : |  :|  | ||: :::|         |  |  ||
Db        422 HTFTTNGTHR-----------VIARYSG--AEGISPSVSLQYPVSVTEAPAADVATTITV 468

Qy        272 DQAANAAESATKDQLTQEAFKNPENQK--------------VNIDANGNA 307
              |  |: |: :    :|  |  :| : :             | :|||| |
Db        469 DPIASTAKGS---PVTLTARLDPADARGTVQFKLGDVLLGGPVRVDANGVA 516
```

RESULT 13
ABB59423
ID ABB59423 standard; Protein; 1472 AA.
XX
AC ABB59423;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 5061.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL03526.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 5061; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1472 AA;

Query Match 5.0%; Score 140.5; DB 22; Length 1472;
Best Local Similarity 21.0%; Pred. No. 0.092;
Matches 121; Conservative 75; Mismatches 240; Indels 139; Gaps 23;

```
Qy          8 NNVNGNHLIPPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIP 67
              | |:|      | |  : |: :|  |  | :   |         | |  | :       |
Db        343 NVVSGTGASVTAVPAAAGTEISAPPGLEHQLVQQG--------SHLEESSSSG------P 388

Qy         68 GLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTHAAIGEKN 127
                | |: |: : :   |    | :    |    || |          |  || |
Db        389 AAVTPPATLSGSATTPLLQYSAAVSNPPPQLQSQGTQSGAGT--------GASAAAGGGA 440

Qy        128 GLEVS--VTLSPQEWSS----------------LQSIDTEGKNRFVFTGGRGGSGHPMV 168
              |   |  |: ||  :||                ||   |  |              |
Db        441 GSTPSSFVSASPDTFSSAASAAATLVHQAQKQQQLQQQTTPIK-------------PSA 486

Qy        169 TVASDIAEARTRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSV 228
              |:: : ::    : ::       |  |  |  :|   | |    : | : : || |:  |
Db        487 TLSVEQSQYFNSLASQ-------GVSPGSVPVQSAPAGYA---QNPVAAYSQTS-TSVGV 535

Qy        229 RSDPKFWVSVGAIAAGLAGLAATG--------IAQALALTPEPDDPTTTDPDQAANAAES 280
                | : :|   |:| |  | |         | :|    | |
Db        536 SQYPNTYANV--FASGTAAGAGTAEQSQQQPQIRRARVKLPPP----------------- 576

Qy        281 ATKDQLTQEAFKNPENQKVN----IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQQA 336
                ::   | : | :  :|    :|     |:  |   ||  | : ::      :  |
Db        577 ---SKIPASAVEMPGDNALNNIGYLDVQFGALDFG--TDDGFEPLPEKVGSGFSIDGQ-- 629

Qy        337 VESNAQAQQRYEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQT 396
                | ||: :|  :: |:: |::  :  || |: |      || |:   |     :|
Db        630 -----QQQQQPDDYQSKSQQQQQVT--LAAGLQSSQISDALNAAGYTS---RSTSQQQQG 679

Qy        397 TTTTTHTVVQQ--------QTGG---IPQHKVALMPQERRRFSDRRDSQGSVASTHWSDS 445
               ::  :  : |         ||||      |:        :        : |  :|: :::
```

```
Db        680 VSSAVNATIDQLTKSDPYGQTGGSGNAYQNAYQSSGASKTASGFPTTAPGGYSSSTYANV 739

Qy        446 SSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSGPVTGRLIGTPGQG 505

Db        740 QSSVANSYQQQG-----YGSYQPSS--YQQQAGSGAQ------SGTGAVSGG-GGTATQN 785

Qy        506 IQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAA 540

Db        786 IPVGGSSSQNSTSGNASSAYLTSGYSTPQSAYQSS 820
```

RESULT 14
ABP43746
ID ABP43746 standard; Protein; 985 AA.
XX
AC ABP43746;
XX
DT 26-FEB-2003 (first entry)
XX
DE 2 PUMH2 protein.
XX
KW Neuroprotective; immunomodulator; cancer; chromosome 2;
KW cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnerary.
XX
OS Homo sapiens.
XX
PN WO200231111-A2.
XX
PD 18-APR-2002.
XX
PF 11-OCT-2001; 2001WO-US27760.
XX
PR 12-OCT-2000; 2000US-0687527.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-426278/45.
DR N-PSDB; ABQ60990.
XX
PT New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation -
XX
PS Claim 20; SEQ ID # 649; 357pp + sequence listing; English.
XX
CC The invention relates to 446 newly isolated polynucleotide sequences.
CC The activity of polynucleotides of the invention may be described as,
CC vulnerary, neuroprotective, immunomodulator, cytostatic and
CC anti-inflammatory. Compositions comprising nucleic acids of the invention
CC are useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records
CC ABP43544-ABP43989 represent polypeptides encoded by polynucleotides of
CC the invention.
CC NOTE: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 985 AA;

Query Match 5.0%; Score 139.5; DB 23; Length 985;
Best Local Similarity 22.4%; Pred. No. 0.063;
Matches 132; Conservative 65; Mismatches 221; Indels 171; Gaps 30;

```
Qy         39 SSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL 98

Db         85 SESGGLGVSMVEYVLSSSPADKLDSRFRKG------NFGTRDAET---DGPEKGDQKGK- 134

Qy         99 DILNTQIGPSAFRVE-----VQADGTHAAI---GEKNGLEVSV-----TLSPQEWSSLQS 145

Db        135 --------ASPFEEDQNRDLKQGDDDDSKINGRGLPNGMDADCKDFNRTPGSRQASPTEV 186

Qy        146 IDTEGKNRFVFTGGRGGSG--------HPMVTVASD----IAEARTRI-LAKLDPDNHGG 192

Db        187 VERLGPN----TNPSEGLGPLPNPTANKPLVEEFSNPETQNLDAMEQVGLESLQFDYPGN 242

Qy        193 RQPKDVDTRSVGVGSASGIDDGVVSETHTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATG 252

Db        243 QVPMDSSGATVGLFDYNS-QQQLFQRTNALTVQQLTAAQQQQYALAAAQQPHIAGVFSAG 301

Qy        253 IAQALALTPEP----DDPTTTDPDQAANAAESATKDQLTQEAFKNPENQKVNIDANGNAI 308

Db        302 LAPA-AFVPNPYIISAAPPGTDPYTAAGLAAAAT---LAGPAVVPPQYY---------GV 348

Qy        309 PSGELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRY-------------EDQHARRQ 355

Db        349 PWGVYPANLFQQ--QAAAAANNTASQQAASQAQPGQQQVLRAGAGQRPLTPNQGQQGQQA 406

Qy        356 EELQLSS------GIGYGLSSAL-----------------IVAG-----GIGAGV----- 382

Db        407 ESLAAAAAANPTLAFGQGLATGMPGYQVLAPTAYYDQTGALVVGPGARTGLGAPVRLMAP 466

Qy        383 TTAL--HRRNQPAEQTTTTTTHTVVQQQTGGI--------PQHKVALMPQERRRFSDRRD 432

Db        467 TPVLISSAAAQAAAAAAAGGTASSLTGSTNGLFRPIGTQPPQ-------QQQQQPSTNLQ 519

Qy        433 SQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSGSG 492

Db        520 SNSFYGSSSLTNSSQS-----------SSLFSHGP---------GQPG-STSLGF-GSG 556

Qy        493 PVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGETAVSSVNAAP 541

Db        557 NSLGAAIGSALSGFGSS-------------GGLTNGSGRYIS---AAP 588
```

RESULT 15
AAB23860
ID AAB23860 standard; Protein; 2411 AA.
XX
AC AAB23860;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-618897/59.
DR N-PSDB; AAA92499.

XX
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -
XX
PS Claim 1; Fig 24; 275pp; English.
XX
CC The present sequence represents a Haemophilus influenzae adhesin
CC (Hia) protein from the type c Haemophilus influenzae strain API.
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
SQ Sequence 2411 AA;

Query Match 5.0%; Score 139.5; DB 21; Length 2411;
Best Local Similarity 21.5%; Pred. No. 0.22;
Matches 125; Conservative 83; Mismatches 228; Indels 145; Gaps 28;

```
Qy          6 LGNNVNGNH----LIPPAPPLPSQTDGAARG---GTGHLISST--GALGSRSLFSPLRNS 56
              ||   || :    :|       :  :||| |     |: || |  |         : ::::
Db       1161 LGGTANGRNDTGTVINKDGLTITLANGAAAGTDASNGNTISVTKDGISAGNKEITNVKSA 1220

Qy         57 MADSVDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA 116
             :    |:::  |  | |   || |:|             :  :|:     | :    : :|
Db       1221 LKTYKDTQNTAG-ATQP---AANTAEVA---------KQDLVDLTKPATGAAGNGADAKA 1267

Qy        117 -DGTHAAIGEKNGL-------EVSVTLSPQEW----SSLQSIDTEGKNRFVFTGGRGGSG 164
               | | | :|:  ||       : :     :|:     :   ::  |||    :     :|
Db       1268 PDTTAATVGDLRGLGWVLSAKKTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNG 1327

Qy        165 HPMVTVASDIAEAR--------TRILAKLDPDNHGGRQPKDVDTRSVGVGSASGIDDGVV 216
               ||:  |:|||:         |    ||  ||  |     ||  : |   | |  : |
Db       1328 KHTVTI--DVAEAKVGDGLEKDTDGKIKLKVDNTDGNNLLTVDA-TKGASVAKGEFNAVT 1384

Qy        217 SETHTST-TNSSVR---------------SDPKFWVSVGAIAAGLAGLAATGIAQALALT 260
              ::  |:  ||:: |               :| |   :|| :|  :   ||| :
Db       1385 TDATTAQGTNANERGKVVVKGSNGATATETDKKKVATVGDVAKAIND-AATFV------K 1437

Qy        261 PEPDDPTTTDPDQAANAAESATK--DQLTQEAFKNPENQKVNID--------ANGNAIPS 310
              | ||  | |     : |  | |  | || :| |    | ||  |         ||  :: |
Db       1438 VENDDSATIDDSPTDDGANDALKAGDTLTLKAGK---NLKVKRDGKNITFALANDLSVKS 1494

Qy        311 GELXDDIVEQIAQQAKEAGEVARQQAVESNAQAQQRYEDQHARRQEELQLSSGIGYGLSS 370
              : | :           |    :  : |: :     :|         : | :||   |:
Db       1495 ATVSDKL---------SLGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTD 1544

Qy        371 ALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHTVVQQQTGGIPQHKVALMPQERRRFSDR 430
              |     : :| || |                          ||       :   |::| :
Db       1545 TL-----LNSGATTNL----------------------GG-----NGITDNEKKRAASV 1571

Qy        431 RDSQGSVASTHWSDSSSEVVNPYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFSG 490
              :|               |:|    | | : : ||:   |:| | ||    |  :   ||
Db       1572 KD---------------VLNAGWNVRGVKPA-SANNQVENI-DFVAT---YDTVDFVSG 1610

Qy        491 SGPVTGRLIGTPGQGIQSTYALLANSGGLRLGMGGLTSGGE 531
                |   : :   | ::   : | :  ::   | | :| |
Db       1611 DKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHNGKLFTGKE 1651
```

Search completed: December 16, 2003, 08:52:53
Job time : 56.0171 secs